Quantile-based Smooth Transition Value at Risk Estimation

Stefan Hubner[1] and Pavel Čížek[2]

[1] *Department of Economics, Oxford University**

[2] *CentER and Department of Econometrics and Operations Research, Tilburg University*

April 3, 2019

Abstract *Value at Risk models are concerned with the estimation of conditional quantiles of a time series. Formally these quantities are a function of conditional volatility and the respective quantile of the innovation distribution. The former is often subject to asymmetric dynamic behaviour, e.g. with respect to past shocks. In this paper we propose a model in which conditional quantiles follow a generalised autoregressive process governed by two parameter regimes with their weights determined by a smooth transition function. We develop a two step estimation procedure based on a sieve estimator, approximating conditional volatility using composite quantile regression, which is then used in the generalised autoregressive conditional quantile estimation. We show the estimator is consistent and asymptotically normal and complement the results with a simulation study. In our empirical application we consider daily returns of the German equity index (DAX) and the USD/GBP exchange rate. While only the latter follows a two regime model, we find that our model performs well in terms of out-of-sample prediction in both cases.*

JEL Codes: *C13, C15, C22, C53*

Keywords: *CAViaR, Composite Quantile Regression, Conditional Quantiles, GARCH, Regime Switching, Smooth Transition, Sieve Estimation*

1 Introduction

With increasing regulatory efforts and new standards for determining capital requirements for financial institutions and the associated importance of effective risk management, methods for estimating conditional volatilities and Value at Risk have been getting significantly

*Corresponding Author: Stefan Hubner (stefan.hubner@economics.ox.ac.uk)

more attention. While a vast amount of models for conditional variance has been developed with Engle (1982) and Bollerslev (1986) leading the way by the Autoregressive and Generalised Autoregressive Conditional Heteroscedasticity (ARCH and GARCH) models, only very few models exist for directly estimating conditional quantiles. The main ones include the conditional quantile ARCH model (Koenker & Zhao, 1996) and the Conditional Autoregressive Value at Risk (CAViaR) model by Engle & Manganelli (2004), which can be interpreted as the conditional quantile analogue of the GARCH model. For a comprehensive discussion of different Value at Risk estimators and their respective merits see Xiao, Guo & Lam (2015).

Although there is a link between conditional volatility and conditional quantiles, which allows the construction of a Value at Risk estimate based on a conditional volatility estimate using a parametrically specified distribution of the error terms (see e.g. Alexander & Leigh (1997), Frey & McNeil (1998), Richardson, Boudoukh & Whitelaw (1998), Adesi, Giannopoulos & Vosper (1999), Gourieroux, Laurent & Scaillet (2000) and Scaillet (2004)), specifying a wrong error distribution can adversely influence the estimates and interpretation via two separate channels. First, the Maximum Likelihood based approaches, which are usually employed for GARCH estimation, directly depend on the correct specification of the innovation distribution. Second, in order to construct the $100\tau\%$-Value at Risk based on these estimates the τ^{th} quantile of the innovations is required.[1] The second channel is particularly harmful under a parametric distributional assumption, especially if the interest lies in the tail estimation as is the case for the Value at Risk. Thus it is preferable to estimate the conditional quantile directly without requiring an assumption about the shape of the error distribution.

Further, it is considered a stylised fact in financial time series that dynamics with respect to positive and negative news are different. In particular there is empirical evidence indicating that volatility is often high after a negative shock, compared to a positive one of equal magnitude (Black, 1976). Theoretically this can be justified by the leverage effect and volatility feedbacks (Andersen & Bollerslev, 2006) or behavioural factors such as loss aversion (McQueen & Vorkink, 2004). Alternatively, time series may be subject to cyclicality which is also not captured by linear models (Tong & Lim, 1980). This can for example be a consequence of business cycles. In any of these cases, it is beneficial and will improve the accuracy of forecasts if one allows for such asymmetric dynamic behaviour. A very general approach to modelling of asymmetric responses to past shocks is the smooth transition approach of Terasvirta (1992), in which the data generating process is driven by and moves between two separate regimes and which includes the threshold model (Tong

[1] This will be discussed in more detail in Section 2, once the necessary notation has been introduced.

& Lim, 1980) as a limit case. The study by Gerlach, Chen & Chan (2011) demonstrates that the conditional volatility models with two regimes can often model better the Value at Risk, especially for lower quantiles.

In this paper, we introduce a smooth transition generalised autoregressive conditional quantile model in which we allow conditional quantiles to follow an autoregressive process that also depends on past error terms as in Engle & Manganelli (2004) and Xiao & Koenker (2009). We allow for asymmetric responses by specifying two regimes, each represented by its own parameter vector. The active regime or regime weights are determined by a transition function characterised by location and scale parameters and a transition variable that can be both a lag of the dependent variable or an exogenous variable. Our approach is related to Xiao & Koenker (2009), who provide a method to estimate the CAViaR model without regime switching by employing a three-stage procedure: first estimating an ARCH approximation of the model, followed by a minimum-distance estimation step to calculate conditional volatilities, which are then finally used for the estimation of the CAViaR model's parameters. The model and estimation procedure we propose can be seen as an extension of this to a regime-switching framework. In addition to this, we improve the original CAViaR estimation by merging the authors' first and second steps using composite quantile regression (Zou & Yuan, 2008), which allows us to eliminate the second step by directly estimating global parameters defining conditional volatilities. Conditionally upon the latter, we can then estimate the CAViaR parameters by using standard quantile regression techniques as in Koenker & Bassett (1978). Our empirical results demonstrate that our model fits the behaviour of two financial time series, the German equity index (DAX) and the USD/GBP exchange rate, in terms of its out-of-sample Value at Risk predictions.

Our study is closely related to the literature on regime switching models, which in its most general form is well established in the context of conditional variance estimation; see Li & Li (1996), Gonzales-Rivera (1998), and Anderson, Nam & Vahid (1999), who use a self-exciting threshold, a smooth transition, and an asymmetric non-linear smooth transition specification, respectively. While some simulation-based research has been done on the topic of modelling regime-switching conditional quantiles, such as White, Tae-Hwan & Manganelli (2008) and Huang *et al.* (2009), who allow for asymmetric responses of autoregressive conditional quantiles without providing any theory, models allowing for asymmetric responses of time series to positive and negative shocks are rather limited in the quantile regression framework, compared to its conditional variance counterpart. Although Engle & Manganelli (2004) propose an asymmetric version of the CAViaR model, namely a Glosten, Jagannathan & Runkle (1993) specification (GJR), they only account for the case where the regime switch is represented by a threshold located at zero and also disregard any asymmetric impacts of past conditional quantiles, which has been empirically documented in economic and financial time series, see for example Nam, Pyun & Avard (2001). An extension of this threshold model, which also allows for two regimes with respect to past conditional quantiles, was studied by Gerlach, Chen & Chan (2011), who demonstrate its performance on a range of different stock market indices. In contrast to threshold models, it is well established that a smooth transition approach facilitates a higher degree of flexibility, by parameterising not only the location at which an instantaneous transition from one regime to the other appears, but also allowing the time series to be in a state determined by any given arbitrary combination of the two polar cases.

Besides not requiring a parametric distributional assumption about the innovations, estimating conditional quantiles rather than the conditional variance concurs with several other properties of quantile regression which prove very useful in this context. First, it allows us to specify a linear structure of conditional volatility as in Taylor (1986) and Schwert (1990). While there exists a quantile regression estimation procedure for a quadratic form of conditional variance (Lee & Noh, 2013), we will instead use such a linear structure of conditional volatility because it has proven to be less sensitive to outliers due to the fact that shocks enter the conditional volatility as a linear absolute value rather than in a squared form. It is well established that the latter leads to an over-prediction of future volatility levels in GARCH models (Klaassen, 2002). Another convenient consequence of a linear specification is that it does not require the existence of the 6^{th} moment for the innovation distribution, but only the $(4+\delta)^{\text{th}}$ moment. Second, regime-switching models are highly non-linear and generally relatively difficult to estimate using traditional numerical methods such as maximum likelihood and thus often result in serious convergence issues, especially if outliers are present. This makes convergence sensitive to the initial parameter value and the choice of the transition function. Quantile regression has the advantage of being numerically very stable at the cost of being computationally more complex. Some of these issues are investigated as part of our simulation study; for a rigorous discussion thereof we refer the reader to Chan & McAleer (2003) and references therein.

2 Model Specification

Let u_t be a stochastic process defined on the real line, from which the stationary sample $\{u_t\}_{t=1}^{n}$ is observed. We assume that this process follows the standard conditional volatility model

$$u_t = \sigma_t\left(z_t, \theta_0\right)\varepsilon_t, \tag{1}$$

where innovations $\{\varepsilon_t\}_{t=1}^n$ are i.i.d. distributed with mean zero and finite variance according to a right-continuous distribution function $F_\varepsilon(x)$ and conditional volatility $\sigma_t : \mathcal{F}_{t-1} \times \Theta_2 \longrightarrow \mathbb{R}_+$ with $\mathcal{F}_{t-1}$ denoting the σ-algebra generated by the process $\{u_s\}_{s=-\infty}^{t-1}$ and Θ_2 denoting the parameter space. Finally, z_t represents the past observations that enter the conditional volatility function and that are assumed to be independent of ε_t.

Instead of the frequently used quadratic specification,[2] in our proposed conditional quantile model we use the following absolute value alternative of the GARCH(p,q) model with $z_t = (\sigma_{t-1}, \ldots, \sigma_{t-p}, |u_{t-1}|, \ldots, |u_{t-q}|)^T$ and parameters $\theta = (\beta_0, \beta_1, \ldots, \beta_p, \gamma_1, \ldots, \gamma_q)^T$:

$$\sigma_t(z_t, \theta) = \beta_0 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}(z_t, \theta) + \sum_{j=1}^{q} \gamma_j |u_{t-j}|. \tag{2}$$

To introduce multiple regimes, we assume that the true conditional volatility process follows a general two regime specification

$$\sigma_t\left(z_t, \theta^{I}, \theta^{II}, \zeta\right) = G\left(\xi_t\left(z_t\right), \zeta, \eta\right) \sigma_t\left(z_t, \theta^{I}\right) + [1 - G\left(\xi_t\left(z_t\right), \zeta, \eta\right)]\sigma_t\left(z_t, \theta^{II}\right), \tag{3}$$

in which each regime is allowed to have different dynamics characterised by regime-specific parameter vectors θ^{I} and θ^{II}, respectively.[3] The parameters θ^{I} and θ^{II} are restricted to be positive to ensure positivity of both conditional volatility processes.[4]

The weight of the active regime in the convex combination of the two regimes in equation (3) is determined by the transition function $G : \mathbb{R}^2 \times \mathbb{R}_+ \longrightarrow [0,1]$, which depends on the transition variable modelled as a pre-specified function $\xi : \mathcal{F}_{t-1} \longrightarrow \mathbb{R}$ of past observations and parameterised by location parameter $\zeta \in \mathbb{R}$ and scale parameter $\eta \in \mathbb{R}_+$. We restrict the function ξ to be time-homogeneous, and as it is assumed to be known, it is referred to as the transition variable $\xi_t := \xi(z_t)$. One example of ξ_t, in the case of daily data, could be the last week's average returns $\xi(z_t) = \frac{1}{5}\sum_{j=1}^{5} u_{t-j}$, where 5 is the typical number of weekly trading days for a financial instrument. Other examples and the selection of the transition variable are discussed in Section 6.

Assumption 1. *The transition function satisfies the following properties:*

$$\lim_{\xi \to -\infty} G(\xi, \zeta, \eta) \to 0 \ \textit{and} \ \lim_{\xi \to +\infty} G(\xi, \zeta, \eta) \to 1,$$

it is monotone, measurable, and Lipschitz. Further, $\partial^d G/\partial(\zeta,\eta)^d$ *exists almost everywhere for* $d = 1$ *and* 2*, is bounded, and is Lipschitz with respect to* ζ *and* η*. In addition to this,* $\partial G/\partial(\zeta,\eta)$ *is monotone or Lipschitz in* ξ*.*

---

[2]The quadratic GARCH(p,q) specification is $\sigma_t(z_t, \theta) = \left(\beta_0 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2(z_t^2, \theta) + \sum_{j=1}^{q} \gamma_j u_{t-j}^2\right)^{1/2}$.

[3]This specification does not represent an extension of Anderson, Nam & Vahid (1999), since the volatility on the right hand side of (3) depends only on θ^{I} and θ^{II}, respectively. We demonstrate the usefulness of this specification in Section 6 and discuss extensions to a fully general setting in Section 5.

[4]Technically, strict positivity only has to hold for some parameters (always including β_0^r for both $r \in \{I, II\}$) and it is allowed that a strict subset of the parameters is non-negative. However, the asymptotic properties we derive are only valid in the interior of the parameter space. We discuss this in more detail in Section 4.

Standard choices for the transition function include:

(i) The logistic function as used in the Logistic Smooth Transition Autoregressive (LSTAR) model by Terasvirta (1992) with $G_{\text{logistic}} : \mathbb{R}^2 \times \mathbb{R}_+ \longrightarrow [0,1]$:

$$G_{\text{logistic}}(\xi, \zeta, \eta) = \left(1 + \exp(-\eta^{-1}(\xi - \zeta))\right)^{-1},$$

(ii) the scale-invariant indicator function $G_{\text{threshold}} : \mathbb{R}^2 \longrightarrow \{0, 1\}$, which reduces the model to the threshold version as in Li & Li (1996):

$$G_{\text{threshold}}(\xi, \zeta) = \mathbb{1}\{\xi \geqslant \zeta\},$$

(iii) and a bounded linear function $G_{\text{linear}} : \mathbb{R}^2 \times \mathbb{R}_+ \longrightarrow [0,1]$ with location ζ centred between two cut-off points:

$$G_{\text{linear}}(\xi, \zeta, \eta) = \left(\eta^{-1}\left(\xi - \zeta + \frac{\eta}{2}\right)\right)\mathbb{1}\left\{\xi \in \left[\zeta - \frac{\eta}{2}, \zeta + \frac{\eta}{2}\right)\right\} + \mathbb{1}\left\{\xi \in [\zeta + \frac{\eta}{2}, \infty)\right\}.$$

Our theoretical results are based on the class of transition functions defined by Assumption 1.[5] For notational convenience, we will stack the transition parameters to the vector $\boldsymbol{\zeta} = (\zeta, \eta)^T$ and abbreviate the transition function as $G_t(\boldsymbol{\zeta}) = G(\xi_t, \zeta, \eta)$.

Having defined the ANST-GARCH model (2)–(3), the shift to the quantile specification is straightforward. The τ^{th} conditional quantile of u_t is defined by

$$Q_{u_t}(\tau|\mathcal{F}_{t-1}) := \inf\left\{x \in \mathbb{R} : F_{u_t|\mathcal{F}_{t-1}}(x) \geqslant \tau\right\}$$

with $F_{u_t|\mathcal{F}_{t-1}}$ being the conditional distribution function of u_t given all past observations, $F_{u_t|\mathcal{F}_{t-1}}(x) = \mathbf{P}(u_t \leqslant x|\mathcal{F}_{t-1})$. For the model defined in equation (1), it follows that

$$\tau = \mathbf{P}\left(u_t \leqslant Q_{u_t}(\tau|\mathcal{F}_{t-1})\right) = \mathbf{P}\left(\sigma_t\varepsilon_t \leqslant Q_{u_t}(\tau|\mathcal{F}_{t-1})\right) = F_\varepsilon\left(\sigma_t^{-1}Q_{u_t}(\tau|\mathcal{F}_{t-1})\right)$$

so that we obtain

$$Q_{u_t}(\tau|\mathcal{F}_{t-1}) = \sigma_t F_\varepsilon^{-1}(\tau). \tag{4}$$

Using this result and multiplying the ANST-GARCH model from equation (3) by $F_\varepsilon^{-1}(\tau)$, for $r \in \{\text{I}, \text{II}\}$ the final asymmetric non-linear smooth transition generalised autoregressive conditional quantile model (ANST-GACQ) can be written as

$$Q_{u_t}(\tau|\mathcal{F}_{t-1}) = G_t(\boldsymbol{\zeta})\, Q^{\text{I}}_{u_t}(\tau|\mathcal{F}_{t-1}) + (1 - G_t(\boldsymbol{\zeta}))Q^{\text{II}}_{u_t}(\tau|\mathcal{F}_{t-1}) \tag{5}$$

$$Q^r_{u_t}(\tau|\mathcal{F}_{t-1}) = \beta^r_0(\tau) + \sum_{i=1}^{p}\beta^r_i Q^r_{u_{t-i}}(\tau|\mathcal{F}_{t-i-1}) + \sum_{j=1}^{q}\gamma^r_j(\tau)\,|u_{t-j}| = \theta^r(\tau)^T\boldsymbol{z}_t, \tag{6}$$

where $\theta^r(\tau) = \left(\beta^r_0(\tau), \beta^r_1(\tau), \ldots, \beta^r_p(\tau), \gamma^r_1(\tau), \ldots, \gamma^r_q(\tau)\right)^T$ and the parameters $\beta^r_i(\tau) := \beta^r_i F_\varepsilon^{-1}(\tau)$ and $\gamma^r_j(\tau) := \gamma^r_j F_\varepsilon^{-1}(\tau)$ for $i \in \mathcal{I}_{1,p}$, $j \in \mathcal{I}_{1,q}$, and $r \in \{\text{I}, \text{II}\}$.[6] These parameters $\beta^r_i(\tau)$, $\gamma^r_j(\tau)$, and thus $\theta^r(\tau)$ are local in the sense that they depend on quantile τ, whereas β^r_i and θ^r are

[5]While there is no doubt that other functions also satisfy Assumption 1, for the empirical part of the study we will restrict the set of transition functions to $\{G_{\text{logistic}}, G_{\text{linear}}\}$. Additionally, we will empirically evaluate data generating processes that follow the limit case with $G_{\text{threshold}}$.

[6]We defined the general index set running from $a \in \mathbb{N}$ to $b \in \mathbb{N}$ as $\mathcal{I}_{a,b} := (a, \ldots, b) \subseteq \mathbb{N}$.

global coefficients independent of quantile τ. The transition parameters ζ are global as well. Global coefficients are not directly identified by the quantile regression. One can however identify them by combining several local ones using composite quantile regression based on K quantiles $\tau_1, \dots, \tau_K \in (0,1)$ since the $2(p+q+1)$ local parameters $\theta^r(\tau)$, $r \in \{I, II\}$, at quantile τ_k are determined by $2(p+q+1)$ global parameters θ^r, $r \in \{I, II\}$, and scalar $F_\varepsilon^{-1}(\tau_k)$; see Section 3. Since $\beta_i^r(\tau) := \beta_i^r F_\varepsilon^{-1}(\tau)$, parameters β_i^r and $F_\varepsilon^{-1}(\tau)$ cannot be separately identified though without some scale normalisation such as $F_\varepsilon^{-1}(\tau_1) = \Phi^{-1}(\tau_1)$, where Φ denotes the standard normal distribution function, for instance. As usual in the smooth transition models, the identification of parameter vectors θ^r, $r \in \{I, II\}$, also presumes the existence of two distinct regimes (otherwise, only their convex combination is identified).

Inversion

The estimation procedure which we will introduce in Section 3 and study in Section 4 is based on invertibility of both GARCH regimes. According to the structure of our model as defined in equation (3) each regime's volatility $\sigma_t\,(z_t, \theta^r)$ does not depend on $\theta^{r'}$ for $r \neq r'$ through lags of σ_t. This implies that both GARCH regimes are required to be additively separable in their respective ARCH and GARCH parts, as discussed in Mele & Fornari (1997).

Assumption 2. *Let* $A^r(L) := 1 - \sum_{i=1}^{p} \beta_i^r L^i$ *and* $B^r(L) := \sum_{i=0}^{q-1} \gamma_i^r L^i$ *where* L *denotes the lag-operator* L*, such that* $u_{t-1} = L u_t$ *for any* $t \in \mathcal{I}_{1,n}$. *The polynomials* $A^r(L)$ *and* $B^r(L)$ *have no common factors and their roots lie outside the unit disc of the complex plane: for* $r \in \{I, II\}$ *and* $|\phi| \leqslant 1$*, it holds that* $A^r(\phi) \neq 0$ *and* $B^r(\phi) \neq 0$.

Hence both GARCH(p,q) regimes defined in equation (2) can be inverted separately,

$$A^r(L)\sigma_t^r = B^r(L)\,|u_t| \iff \sigma_t^r = A^{r^{-1}}(L)B^r(L)\,|u_t| = \alpha_0^r + \sum_{j=1}^{\infty} \alpha_j^r\,|u_{t-j-1}|, \tag{7}$$

where $r \in \{I, II\}$ and the coefficients α_j for $j \in \mathcal{I}_{1,m}$ decrease at a geometric rate, that is, there exist constants $b < 1$ and c such that $|\alpha_j| < cb^j$.

While this is in line with various GARCH extensions to two regimes (Glosten, Jagannathan & Runkle (1993), Gonzales-Rivera (1998), Rabemananjara & Zakoian (1993)) and is equivalent for the threshold model where $G_t(\zeta) \in \{0, 1\}$, it differs from the specification of Anderson, Nam & Vahid (1999). The reason for this is that the latter imposes a smooth transition between two DGP's and the transition acts upon the individual coefficients, whereas our model impose a smooth transition between two volatility processes and the reality is a convex combination of their outcomes.

In the following sections we discuss estimation and asymptotic properties of the additively separable model defined in (3). Further, we discuss how the proposed estimation procedure can be applied to the Anderson, Nam & Vahid (1999) specification

$$\sigma_t\left(z_t, \theta^{I}, \theta^{II}, \zeta\right) = G_t(\zeta)\, \sigma(z_t, \theta^{I}, \theta^{II})) + (1 - G_t(\zeta))\sigma(z_t, \theta^{I}, \theta^{II}), \tag{8}$$

by discussing an extension in Section 5. In addition to this, we develop a test of the additively separable specification against the one in (8). We show that our model is empirically relevant in two different applications.

3 Estimation Procedure

The estimation of the CAViaR model specified in Section 2 is complicated due to the dependence of conditional quantiles on past conditional quantiles in equation (6).To address this, we propose a two-step estimation procedure that is related to the three-stage sieve approximation idea of Xiao & Koenker (2009). In contrast to their single regime version, our model requires the estimation of parameter vectors for both regimes as well as the location and scale parameters of the transition function.

The proposed estimation procedure consists of two steps: first, we approximate the conditional volatility process defined in equation (2) by an ARCH(∞)-approximation to deal with the dependence of unknown conditional quantiles; second, after obtaining approximations of the conditional volatilities from the first step, the model structure (4) and formulation (6) is used to estimate the CAViaR parameters and the transition parameters by the quantile regression. What further complicates estimation is the fact that, although the transition function is assumed to be known *a priori*, the objective function is not necessarily convex in all of its parameters. In order to estimate the parameters of the transition function, we therefore have to combine linear quantile regression with a grid search in both steps.

The model defined in (5) and (6) can be estimated using the objective function

$$\min_{\theta \in \Theta_2^{\tau}} n^{-1} \sum_{t=1}^{n} \rho_\tau \left(u_t - {\theta^{I}}^{\top} z_t(\theta) G_t(\zeta) - {\theta^{II}}^{\top} z_t(\theta)(1 - G_t(\zeta)) \right),$$

where function $\rho_\tau(u) = u\left(\tau - \mathbb{1}_{\{u<0\}}\right)$ denotes the quantile loss function, $z_t(\theta) = (\sigma_{t-1}(\theta), \ldots, \sigma_{t-p}(\theta), |u_{t-1}|, \ldots, |u_{t-q}|)^{\top}$, and $\sigma_t(\theta)$ has the structure defined in (2). The estimation of conditional quantiles would thus be a linear programming exercise, if it were not for the dependence on the latent conditional volatility process σ_t, which in turn dynamically depends on the parameters θ to be estimated. To tackle this issue, a two-step procedure is used.

In the first step, each regime's GARCH(p,q) process in equation (3) is inverted to ARCH(∞) and estimated using an ARCH(m) representation, $m \in \mathbb{N}$, in order to find a sieve approximation of $\sigma_t := \sigma_t(\boldsymbol{z}_t, \theta)$. Consequently, each conditional volatility regime defined in equation (7) can be approximated by an ARCH(m) process up to a reminder term $\mathcal{O}_p(b^m)$ and the ANST-GACQ model can thus be rewritten as

$$Q_{u_t}(\tau|\mathcal{F}_{t-1}) \approx \left(\alpha_0^{\mathrm{I}}(\tau) + \sum_{j=1}^{m} \alpha_j^{\mathrm{I}}(\tau)\,|u_{t-j}|\right) G_t(\zeta) + \left(\alpha_0^{\mathrm{II}}(\tau) + \sum_{j=1}^{m} \alpha_j^{\mathrm{II}}(\tau)\,|u_{t-j}|\right)(1 - G_t(\zeta))$$

with $\alpha_j^r(\tau) := \alpha_j^r F_\varepsilon^{-1}(\tau)$ for all $j \in \mathcal{I}_{0,m}$ and $r \in \{\mathrm{I}, \mathrm{II}\}$. In order to estimate the conditional volatility process $\{\sigma_t\}$, we need to identify and estimate $\boldsymbol{\alpha}^r = (\alpha_1^r, \ldots, \alpha_m^r)^\mathsf{T}$ separately from $F_\varepsilon^{-1}(\tau)$. Moreover, the estimation of the transition parameters ζ is traditionally rather difficult and becomes very imprecise at more extreme quantiles. For this reason, we will not only estimate single conditional quantiles but exploit information from a range of quantiles[7] $\tau_1, \ldots, \tau_K \in (0,1)$ and employ composite quantile regression (Zou & Yuan, 2008) by minimising the following objective function:

$$\widehat{\mathfrak{a}}_n = \arg\min_{\mathfrak{a} \in \Theta_1} \sum_{k=1}^{K} \sum_{t=m+1}^{n} \rho_{\tau_k}\left(u_t - q_k \boldsymbol{\alpha}^{\mathrm{I}^\mathsf{T}} \boldsymbol{z}_t^m G_t(\zeta) - q_k \boldsymbol{\alpha}^{\mathrm{II}^\mathsf{T}} \boldsymbol{z}_t^m (1 - G_t(\zeta))\right), \tag{9}$$

where $\boldsymbol{z}_t^m = (1, |u_{t-1}|, \ldots, |u_{t-m}|)^\mathsf{T}$, $\boldsymbol{\alpha}^r = (\alpha_0^r, \ldots, \alpha_m^r)^\mathsf{T}$ for $r \in \{\mathrm{I}, \mathrm{II}\}$, $\mathbf{q} = (q_1, \ldots, q_K)^\mathsf{T}$ with $q_k = F_\varepsilon^{-1}(\tau_k)$ for all $k \in \mathcal{I}_{1,K}$, and $\mathfrak{a} = \left[\boldsymbol{\alpha}^{\mathrm{I}^\mathsf{T}}, \boldsymbol{\alpha}^{\mathrm{II}^\mathsf{T}}, \mathbf{q}^\mathsf{T}, \zeta^\mathsf{T}\right]^\mathsf{T} \in \boldsymbol{\Theta}_1$ that is assumed to be a compact subset of $\mathbb{R}_+^{2(m+1)} \times \mathbb{R}^{K+2}$. Similarly to the notation in (5) and (6), the global parameters describing the evolution of conditional volatility are denoted $\boldsymbol{\alpha} = \left[\boldsymbol{\alpha}^{\mathrm{I}^\mathsf{T}}, \boldsymbol{\alpha}^{\mathrm{II}^\mathsf{T}}, \zeta^\mathsf{T}\right]^\mathsf{T}$ and its estimate $\widehat{\boldsymbol{\alpha}}_n$ (which contains all elements of $\widehat{\mathfrak{a}}_n$ except $\widehat{\mathbf{q}}_n$). As discussed in Section 2, one element of $\mathbf{q}$ has to be fixed in (9) to achieve identification, for example, $q_1 = \Phi^{-1}(\tau_1)$ for $\tau_1 \neq 0.5$ (this normalisation has of course no effect on the estimated volatility process). In addition to this, we estimate $\widehat{\boldsymbol{\alpha}}_n$ subject to a non-negativity constraint to ensure that both volatility processes are positive. The advantage of the proposed composite quantile criterion lies in the joint estimation of the conditional volatility parameters in a single step, and in particular, of the parameters of the transition function.

In the second step, we can now use the first-stage estimates and equation (7) to approximate $\widehat{\sigma}_t$ by

$$\sigma_t(\widehat{\boldsymbol{\alpha}}_n) = \left(\widehat{\boldsymbol{\alpha}_n^{\mathrm{I}}}^\mathsf{T} \boldsymbol{z}_t^m\right) G_t\left(\widehat{\zeta}_n\right) + \left(\widehat{\boldsymbol{\alpha}_n^{\mathrm{II}}}^\mathsf{T} \boldsymbol{z}_t^m\right)\left(1 - G_t\left(\widehat{\zeta}_n\right)\right). \tag{10}$$

Defining $\boldsymbol{z}_t(\widehat{\boldsymbol{\alpha}}_n) = (\sigma_{t-1}(\widehat{\boldsymbol{\alpha}}_n), \ldots, \sigma_{t-p}(\widehat{\boldsymbol{\alpha}}_n), |u_{t-1}|, \ldots, |u_{t-q}|)^\mathsf{T}$, we can estimate the CAViaR model according to equation (5) and (6) for a single quantile $\tau \in (0,1)$ by minimising

$$\widehat{\theta}_n(\tau) = \arg\min_{\theta \in \Theta_2^\tau} \sum_{t=t_0}^{n} \rho_\tau\left(u_t - \theta^{\mathrm{I}^\mathsf{T}} \boldsymbol{z}_t(\widehat{\boldsymbol{\alpha}}_n) G_t(\zeta) - \theta^{\mathrm{II}^\mathsf{T}} \boldsymbol{z}_t(\widehat{\boldsymbol{\alpha}}_n)\,(1 - G_t(\zeta))\right), \tag{11}$$

[7] Xiao & Koenker (2009) solve this by first estimating the parameters for each τ and then exploit their structure $\alpha_j^r(\tau) := \alpha_j^r F_\varepsilon^{-1}(\tau)$ in an additional minimum distance estimation step.

where $\theta(\tau) = \left[\theta^{I^T}(\tau), \theta^{II^T}(\tau), \zeta^T(\tau)\right]^T \in \boldsymbol{\Theta}_2^\tau$ and $t_0 = [(m+p) \vee q] + 1$.

Neither of the objective functions (9) and (11) are convex in the scale parameter. In addition to this, the quantile loss function is not differentiable. For this reason, we use a grid search over the space of feasible scale parameters in the first stage (this would also apply to the location parameter if the transition function is not monotonic). Moreover, to minimise the composite quantile criterion (9) for a given value of the scale parameter, we employ a smoothed version of the objective function ρ defined by $\rho^*(u) := \rho(u)$ if $|u| > \delta$ and u^2/δ otherwise with smoothing parameter δ. This approach is commonly used in the literature (see e.g. Huber (1964); Zheng (2011)) and facilitates estimation of the parameters for a given grid point using gradient based methods. In particular, we use sequential quadratic programming to incorporate the inequality constraints on the ARCH(m) parameters. In the second stage, for a given location/scale parameter pair ζ obtained now from a two-dimensional grid search, we can estimate $\theta(\tau)$ using standard quantile autoregression (Koenker & Zhao, 1996). To ensure positivity of the global coefficients, which translates to negativity of local ones for $\tau < 0.5$, we use an interior-point method for inequality constrained quantile regression (Koenker & Ng, 2005). Note that, we do not make use of the estimated quantiles $\widehat{q}_n$ from the first stage, and additionally, we also re-estimate the location and scale parameters, which we denote by $\zeta(\tau)$ in the second stage to indicate the local estimation at one quantile τ. Algorithm A.2.1 summarises the whole procedure as pseudo-code and can be found in Section A.2 of the online appendix. We denote the one- and two-dimensional grids as $\{\eta_1, \ldots, \eta_{k_\eta}\}$ and $\{\zeta_1, \ldots \zeta_{k_\zeta}\} \times \{\eta_1, \ldots, \eta_{k_\eta}\}$ and informally denote the subspace of feasible location and scale pairs by $\mathfrak{Z}$.

The first-stage of Algorithm A.2.1 relies on an auxiliary set of K quantiles $(\tau_1, \ldots, \tau_K)$. Although a higher K and thus more quantiles allow us to recover more information about the distribution of u_t, the parameters θ_j for $j \in \mathcal{I}_{1,2(p+q+1)}$ are not identified at the median due to the model structure $\theta_j(\tau) = F_\varepsilon^{-1}(\tau)\theta_j$. Due to this lack of identification, one would thus introduce extra noise by including quantiles at or close to $\tau = 0.5$ in finite samples. We thus face a bias-variance trade-off in the selection of K. While a data-driven optimal choice of a vector of τ's would be feasible, this goes beyond the scope of this paper; we refer interested readers to Zhao & Xiao (2014). We assume that quantiles $\tau_k \notin (0.5-\delta/2, 0.5+\delta/2)$, $k \in \mathcal{I}_{1,K}$, in (9) and demonstrate the insensitivity of the method to the choice of δ in our simulation study; see Section A.4 of the online appendix.

4 Asymptotic Results

In this section, the two-stage estimation procedure introduced for the proposed ANST-GACQ model is shown to yield consistent and asymptotically normal estimates. To this end, the true parameter values, minimising the corresponding population objective functions, are labelled with subscript 0. All proofs can be found in Section A.1 of the supplementary online appendix. Throughout this section, it is assumed that, in addition to the previously defined assumptions, the following statements hold.

Assumption 3. *The errors* ε_t *are independent and identically distributed with zero median and finite variance* $\sigma^2 = Var(\varepsilon_t) < +\infty$. *The distribution function* $F_\varepsilon(x)$ *has a strictly positive density* $f_\varepsilon(x)$ *at* $F_\varepsilon^{-1}(\tau_k)$ *for all* $k \in \mathcal{I}_{1,K}$, *which is uniformly bounded by a finite constant* M *and is Lipschitz continuous.*

Assumption 4. *The conditional distribution function* $F_{u_t|\mathcal{F}_{t-1}}(x)$ *has a strictly positive density* $f_{u_t|\mathcal{F}_{t-1}}(x)$ *at* $F^{-1}_{u_t|\mathcal{F}_{t-1}}(\tau_k)$ *for all* $k \in \mathcal{I}_{1,K}$, *which is uniformly bounded by a finite constant* M *and is Lipschitz continuous.*

Assumption 5. *There exist small positive constants* $\gamma > 0$ *and* $\delta > 0$ *such that* $\mathbb{E}\,|u_t G_t(\zeta_0)|^{2+\delta} < +\infty$, $\mathbb{E}\,|u_t|^{2+\delta} < +\infty$, *and* $\mathbb{E}\,\|u_t \partial G_t(\zeta_0)/\partial\zeta\|^{2+\delta} < +\infty$, *and additionally,* u_t *is strictly stationary and* β*-mixing with mixing coefficients* $\beta_s = \mathcal{O}(s^{-\max\{2,(2+\delta)/\delta\}-\gamma})$ *as* $s \to \infty$.

Assumption 6. *Let* $a = [\boldsymbol{\alpha}^{I^\top}, \boldsymbol{\alpha}^{II^\top}, q^\top, \zeta^\top]^\top$ *and*

$$\mathbf{x}_{t,k}(a) = \begin{bmatrix} q_k I_{2(p+q+1)} \\ \mathbb{1}\{k=1\}[\boldsymbol{\alpha}^{I^\top}, \boldsymbol{\alpha}^{II^\top}]^\top \\ \vdots \\ \mathbb{1}\{k=K\}[\boldsymbol{\alpha}^{I^\top}, \boldsymbol{\alpha}^{II^\top}]^\top \\ q_k(\boldsymbol{\alpha}^{I} - \boldsymbol{\alpha}^{II})^\top \frac{\partial G_t(\zeta)}{\partial \zeta} \\ q_k(\boldsymbol{\alpha}^{I} - \boldsymbol{\alpha}^{II})^\top \frac{\partial G_t(\zeta)}{\partial \eta} \end{bmatrix} \begin{bmatrix} G_t(\zeta)\, \boldsymbol{z}_t^m \\ (1 - G_t(\zeta))\boldsymbol{z}_t^m \end{bmatrix},$$

see equation (18). *The matrix*

$$\mathbf{D}_{1,m,n}(a) := \mathbb{E}\left[n^{-1}\sum_{t=m}^{n}\left[\sum_{k=1}^{K}\mathbf{x}_{t,k}(a)\mathbf{x}_{t,k}(a)^\top\right]/\sigma_t\right]$$

evaluated at a_0 *has minimum and maximum eigenvalues* $\lambda_{n,\min}$ *and* $\lambda_{n,\max}$ *satisfying* $\lim_{n\to\infty}\inf \lambda_{n,\min} > 0$ *and* $\lim_{n\to\infty}\sup \lambda_{n,\max} < +\infty$. *Additionally, assume that*

$$\mathbb{E}\begin{bmatrix} G_t(\zeta)\,\boldsymbol{z}_t^m \\ (1-G_t(\zeta))\boldsymbol{z}_t^m \\ (G_t(\zeta) - G_t(\zeta_0))\boldsymbol{z}_t^m \end{bmatrix}\begin{bmatrix} G_t(\zeta)\,\boldsymbol{z}_t^m \\ (1-G_t(\zeta))\boldsymbol{z}_t^m \\ (G_t(\zeta) - G_t(\zeta_0))\boldsymbol{z}_t^m \end{bmatrix}^\top$$

has the full rank for any $\zeta \neq \zeta_0$.

Assumption 7. *The number of lags for the ARCH(m)-approximation satisfies* $\log(n)/m \to 0$ *and* $mn^{-\frac{1}{2}} \to 0$*. It holds for the number* K *of quantiles different from 0.5 that* $K > 1$.

Assumption 8. *There exist small positive constants* $\delta_1 > 0$ *and* $\delta_2 > 0$ *such that*

$$\mathbf{P}\left(\max_{1\leqslant t\leqslant n} u_t^2 > n^{\delta_1}\right) \leqslant \exp\left(-n^{\delta_2}\right).$$

Assumptions 2–5 guarantee that the process u_t is stationary and weakly dependent. These assumptions facilitate deriving general results, but it is also possible to find more primitive sufficient conditions that guarantee the stationarity and β-mixing of particular regime switching models using results of Carrasco & Chen (2002) and Meitz & Saikkonen (2008), for instance. We discuss sufficient conditions in Appendix A.3 and show that they are the same in the proposed specification (3) and the general specification (8). For example in the case of the frequently used GARCH(1,1) model with the lagged dependent variable serving as the transition variable, the stationarity and β-mixing properties require $\max\{\beta_1^{I}, \beta_1^{II}\} + \max\{\gamma_1^{I}, \gamma_2^{II}\} < 1$ in models (3) and (8).

Assumption 5 also imposes the moment assumptions on u_t, $u_t G_t(\zeta_0)$, and $u_t \partial G_t(\zeta)/\partial\zeta$ that are required for central limit theorems under weak dependence. Note that a sufficient condition for the existence of the finite $(2+\delta)$ moment of the term $u_t \partial G_t(\zeta)/\partial\zeta$ is the existence of the $(4+\delta)$th moments for u_t and $\mathbb{E}\,|\partial G_t(\zeta)/\partial\zeta| < \infty$. In the case of exogenous switching, that is, if ξ_t is an exogenous time series, the existence of $(2+\delta)$th finite moments of u_t suffices. Further by imposing the full rank assumptions on the matrices appearing in the first and second stage first-order conditions, Assumption 6 provides identification, ensuring that the two regimes have different conditional volatility processes, that the data in the two regimes are not perfectly correlated and that, for the transition functions that have zero slope on subsets of its domain, there are data in both regimes with positive probability. Next, Assumption 7 restricts the rate of the ARCH(m) approximation, ensuring that we have a sufficient number of lags to control the approximation error. In our empirical application we will choose $m = cn^{1/4}$ for some positive constant $c > 0$. We demonstrate the accuracy of the estimates with respect to the choice of c as part of our simulation study; see Section A.4 of the online appendix. Finally, Assumption 8 is a technical regularity condition that is needed for the sieve estimation in the first stage (Xiao & Koenker, 2009).

We will now present the asymptotic properties of our estimation procedure. First, we show that the sieve approximation of both regimes' underlying GARCH processes holds and the approximation error due to the m$^{\text{th}}$-order truncation is bounded in probability.

Theorem 1. *Let the parameter vector be defined as* $a = [\alpha^{I^\top}, \alpha^{II^\top}, q^\top, \zeta^\top]^\top$*. Under Assumptions 1–8,* a *is identified and it holds for* $n \to \infty$ *that* $\|\widehat{a}_n - a_0\|^2 = \mathcal{O}_p\left(mn^{-1}\right)$.

The following corollary provides the asymptotic characteristics of the interim sieve estimator for the latent process σ_t defined as a function of a according to equation (10) and as such a preliminary result for the second stage estimator where we use quantile regression with $z_t(\widehat{a}_n) = (\sigma_{t-1}(\widehat{a}_n), \ldots, \sigma_{t-p}(\widehat{a}_n), |u_{t-1}|, \ldots, |u_{t-q}|)^T$ to obtain the final CAViaR parameters $\theta(\tau)$. Before finding the asymptotic distribution of the second stage estimator, Assumption 9 defines several matrices.

Theorem 2. *Let* $s_k = f_\varepsilon(F_\varepsilon^{-1}(\tau_k))$ *and* $S_0 = \sum_{k=1}^{K} s_k q_k^2$*. Also let* $\boldsymbol{\alpha}^\Delta = \boldsymbol{\alpha}^I - \boldsymbol{\alpha}^{II}$*. Under Assumptions 1–8, it then holds for* $n \to \infty$ *that*

$$\sqrt{n}\begin{bmatrix}\widehat{\boldsymbol{\alpha}_n^I}-\boldsymbol{\alpha}_0^I\\ \widehat{\boldsymbol{\alpha}_n^{II}}-\boldsymbol{\alpha}_0^{II}\\ \widehat{\zeta_n}-\zeta_0\end{bmatrix} \approx \mathbf{D}_m^{-1}\frac{1}{S_0\sqrt{n}}\sum_{t=m+1}^{N}\begin{bmatrix}G_t(\zeta)\,\boldsymbol{z}_t^m\\ (1-G_t(\zeta))\boldsymbol{z}_t^m\\ \boldsymbol{\alpha}_0^{\Delta^T}\boldsymbol{z}_t^m\frac{\partial G_t(\zeta)}{\partial \zeta}\end{bmatrix}\sum_{k=1}^{K} q_k\left(\mathbb{1}\left\{u_t \leqslant F^{-1}_{u_t|\mathcal{F}_{t-1}}(\tau_k)\right\}-\tau_k\right)$$

with reminder of stochastic order $o_p\left(m^{\frac{1}{2}}n^{-\frac{1}{2}}\right)$ *and* $\mathbf{D}_m :=$

$$-\mathbb{E}\left[\frac{1}{\sigma_t}\left[\begin{array}{c|c}\begin{bmatrix}G_t(\zeta_0)\\1-G_t(\zeta_0)\end{bmatrix}\begin{bmatrix}G_t(\zeta_0)\\1-G_t(\zeta_0)\end{bmatrix}^T\otimes\boldsymbol{z}_t^m\boldsymbol{z}_t^{mT} & \boldsymbol{z}_t^{mT}\boldsymbol{\alpha}_0^\Delta\left(\begin{bmatrix}G_t(\zeta_0)\\1-G_t(\zeta_0)\end{bmatrix}\otimes\boldsymbol{z}_t^m\right)\frac{\partial G_t(\zeta_0)}{\partial\zeta^T}\\ \hline & (\boldsymbol{z}_t^{mT}\boldsymbol{\alpha}_0^\Delta)^2\frac{\partial G_t(\zeta_0)}{\partial\zeta}\frac{\partial G_t(\zeta_0)}{\partial\zeta^T}\end{array}\right]\right].$$

Assumption 9. *Let* $\Gamma_{\theta,0} = \Gamma_\theta(\theta_0, \boldsymbol{\alpha}_0)$ *and* $\Gamma_{\alpha,m,0} := \Gamma_{\alpha,m}(\theta_0, \boldsymbol{\alpha}_0)$ *exist, where*

$$\Gamma_\theta(\theta,\boldsymbol{\alpha}) := -\frac{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)}{\sigma_\varepsilon}\times$$
$$\mathbb{E}\left[\begin{array}{c|c}\begin{bmatrix}G_t(\zeta)\\1-G_t(\zeta)\end{bmatrix}\begin{bmatrix}G_t(\zeta)\\1-G_t(\zeta)\end{bmatrix}^T\otimes\boldsymbol{z}_t(\boldsymbol{\alpha})\boldsymbol{z}_t(\boldsymbol{\alpha})^T & \theta^{\Delta^T}\boldsymbol{z}_t(\boldsymbol{\alpha})\begin{bmatrix}G_t(\zeta)\\1-G_t(\zeta)\end{bmatrix}\otimes\boldsymbol{z}_t(\boldsymbol{\alpha})\frac{\partial G_t(\zeta)}{\partial\zeta^T}\\ \hline & \left(\theta^{\Delta T}\boldsymbol{z}_t(\boldsymbol{\alpha})\right)^2\frac{\partial G_t(\zeta)}{\partial\zeta}\frac{\partial G_t(\zeta)}{\partial\zeta^T}\end{array}\right]$$

and

$$\Gamma_{\alpha,m}(\theta,\boldsymbol{\alpha}) := -\frac{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)}{\sigma_\varepsilon}\times$$
$$\mathbb{E}\left[\begin{bmatrix}\boldsymbol{z}_t(\boldsymbol{\alpha})G_t(\zeta)\\ \boldsymbol{z}_t(\boldsymbol{\alpha})(1-G_t(\zeta))\\ \theta^{\Delta T}\boldsymbol{z}_t(\boldsymbol{\alpha})\frac{\partial G_t(\zeta)}{\partial\zeta}\end{bmatrix}\theta^T\left[\begin{bmatrix}G_t(\zeta)\\1-G_t(\zeta)\end{bmatrix}\otimes\begin{bmatrix}[L^1,\ldots,L^q]^T\otimes[\boldsymbol{z}_t^{mT},\boldsymbol{\alpha}^{\Delta^T}\boldsymbol{z}_t^m G_t(\zeta_0)\frac{\partial G_t(\zeta_0)}{\partial\zeta^T}]\\ 0_{p+1,m+2}\end{bmatrix}\right]\right],$$

and L *denotes the lag operator. Further, let* $\Gamma_{\theta,0}$ *be non-singular,*

$$\mathbf{M}_{t,m} = \left[\mathbf{I}_{2(p+q+1)}, \left(\sum_{k=1}^{K} s_k q_k^2\right)^{-1}\Gamma_{\alpha,m,0}\mathbf{D}_m^{-1}\right]\begin{bmatrix}G_t(\zeta_0)\,\boldsymbol{z}_t(\boldsymbol{\alpha}_0)\\ (1-G_t(\zeta_0))\boldsymbol{z}_t(\boldsymbol{\alpha}_0)\\ \theta_0^{\Delta^T}\boldsymbol{z}_t(\boldsymbol{\alpha}_0)\frac{\partial G_t(\zeta_0)}{\partial\zeta}\\ G_t(\zeta_0)\,\boldsymbol{z}_t^m\\ (1-G_t(\zeta_0))\boldsymbol{z}_t^m\\ \boldsymbol{\alpha}_0^{\Delta^T}\boldsymbol{z}_t^m\frac{\partial G_t(\zeta_0)}{\partial\zeta}\end{bmatrix},$$

and for a given value of $\tau \in (0,1)$, Ξ^τ *be a* $2(p+q+1) \times 2(p+q+1)$ *matrix with a typical element*

$$\Xi_{i,j} = \begin{cases} \frac{\tau(1-\tau)}{f_\varepsilon(F_\varepsilon^{-1}(\tau))^2} & \text{if } i \leqslant p+q+1 \text{ and } j \leqslant p+q+1 \\ \frac{q_i q_j(\tau_i \wedge \tau_j)(1-\tau_i \vee \tau_j)}{f_\varepsilon(F_\varepsilon^{-1}(\tau_i)) f_\varepsilon(F_\varepsilon^{-1}(\tau_j))} & \text{if } i > p+q+1 \text{ and } j > p+q+1 \\ \frac{q_i(\tau_i \wedge \tau)(1-\tau_i \vee \tau)}{f_\varepsilon(F_\varepsilon^{-1}(\tau_i)) f_\varepsilon(F_\varepsilon^{-1}(\tau))} & \text{otherwise } (i > j \text{ w.l.o.g}). \end{cases}.$$

Then $\lim_{m\to\infty} \mathbb{E}[\mathbf{M}_{t,m}\Xi^\tau \mathbf{M}_{t,m}]$ *is assumed to exist and be finite.*

The following two theorems provide now consistency and asymptotic normality results of the final ANST-GACQ estimator using the preliminary results from the first stage.

Theorem 3 (Second Stage Consistency)**.** *Under Assumptions 1–9, the second-stage estimator is* $\sqrt{n}$*-consistent, that is, for* $n \to \infty$ *and given* $\tau \in (0,1)$

$$\left\| \widehat{\theta}_n(\tau) - \theta_0(\tau) \right\| = \mathcal{O}_p(n^{-\frac{1}{2}}).$$

Theorem 4 (Second Stage Asymptotic Normality)**.** *If Assumptions 1–9 hold, then for* $\theta_0(\tau) \notin \partial\Theta_2^\tau$ *the second-stage estimator* $\widehat{\theta}_n(\tau)$ *is asymptotically normal, that is, for* $n \to \infty$ *and given* $\tau \in (0,1)$

$$\sqrt{n}\left(\widehat{\theta}_n(\tau) - \theta_0(\tau)\right) \rightsquigarrow \mathcal{N}\left(0, \lim_{m\to\infty} \Gamma_{\theta,0}^{-1} \mathbb{E}[\mathbf{M}_t \Xi^\tau \mathbf{M}_t] \Gamma_{\theta,0}^{-1}\right).$$

The asymptotic distribution and variance established in Theorem 4 can be evaluated by using the finite-sample equivalence of the respective expectations, with the exception of the densities in matrix Ξ^τ. An overview of estimation approaches for evaluating Ξ^τ can be found in Koenker (2005), for instance.

5 Extension & Separability Test

We will now show that our estimation procedure can be extended in such a way that it does not require additive separability and Assumption 2 and it accommodates the specification of Anderson, Nam & Vahid (1999). For this, note that using the definition of the lag-polynomials from Assumption 2, the model (3) can be rewritten as

$$\begin{aligned} \sigma_t\left(z_t, \theta^I, \theta^{II}, \zeta\right) &= G_t(\zeta)\left[\beta_0^I + (1 - A^I(L))\sigma_t\left(z_t, \theta^I, \theta^{II}, \zeta\right) + B^I(L)|u_t|\right] \\ &+ (1 - G_t(\zeta))\left[\beta_0^{II} + (1 - A^{II}(L))\sigma_t\left(z_t, \theta^I, \theta^{II}, \zeta\right) + B^{II}(L)|u_t|\right], \end{aligned}$$

where the σ_t on the right-hand side is now allowed to depend both on θ^I and θ^{II}, and

$$\begin{aligned} & \left[G_t(\zeta)\, A^I(L) + (1 - G_t(\zeta))A^{II}(L)\right] \sigma_t\left(z_t, \theta^I, \theta^{II}, \zeta\right) \\ = & \left[G_t(\zeta)\,(\beta_0^I + B^I(L)) + (1 - G_t(\zeta))(\beta_0^{II} + B^{II}(L))\right] |u_t|. \end{aligned}$$

Conditionally on the transition variable ξ_t, one can again impose invertibility of the lagged polynomial $A_t(L) = G_t(\zeta)\, A^{I}(L) + (1 - G_t(\zeta))A^{II}(L)$ to transform the GARCH model to its ARCH(∞) representation (recall that $G_t(\zeta) \in [0,1]$). Contrary to the model under Assumption 2, this polynomial $A_t(L)$ and its inversion will however depend on $\xi(z_t)$ and the ARCH(∞) representation will therefore vary with $\xi_t(z_t)$.

Since the function G is smooth by Assumption 1, the model (5)–(6) can be estimated analogously to Algorithm A.2.1 if the ARCH approximation (9) is replaced by a general ARCH(m) approximation local to the value $\xi(z_t)$ when predicting the volatility using equation (10):

$$\left[\widehat{\boldsymbol{\alpha}}_n^{T}(z_t), \widehat{\mathbf{q}}_n^{T}(z_t)\right]^{T} = \arg\min \sum_{k=1}^{K} \sum_{s=m+1}^{n} \rho_{\tau_k}\left(u_t - q_k(z_t)\boldsymbol{\alpha}^{T}(z_t) z_s^{m}\right) \mathcal{K}_h\left(\xi(z_s) - \xi(z_t)\right),$$

where $\mathcal{K}$ represents a univariate kernel indexed by a suitable bandwidth h; see for example Cai & Xu (2009) for a discussion on kernel and bandwidth selection. Note, that $\xi(z_t)$ is a scalar and is thus not subject to the curse of dimensionality in this non-parametric estimation. Consequently, the asymptotic results presented in Section 4 will apply if the conditions for value m characterising the order of the ARCH approximation are replaced by analogous conditions on m/h. The reason for this is that because of the local estimation, the number of observations available for the local ARCH(m) estimation will be proportional to nh and the rates in the proof of Theorem 1 will become $m/(nh)$. Since they should be negligible with respect to $1/\sqrt{n}$, this would require $\log(n)h/m \to 0$ and $mn^{-1/2}/h \to 0$ as $n \to \infty$.

We will now show how to construct a test between the two specifications. Lemma 1 shows the structure of the lag-polynomials if the GARCH regimes are not invertible separately as in equation (8).

Lemma 1 (Structure of G_t-Dependent Polynomial)**.** *Let* $\bar{\mathcal{P}}(L^0) = G_t(\zeta)\, c_1 + (1 - G_t(\zeta))c_2$ *for generic*[8] *coefficients* c_1 *and* c_2. *Further let* $\bar{\mathcal{P}}(L^{m+1}) = \bar{\mathcal{P}}(L^m)L\bar{\mathcal{P}}(L^m)$. *Then for any* $m \in \mathbb{N}$ *the specification from equation* (8) *can be represented as*

$$\sigma_t = \bar{\mathcal{P}}(L^m)\sigma_{t-m-1} + \sum_{k=0}^{m} \bar{\mathcal{P}}(L^k)|u_{t-k-1}|, \tag{16}$$

with the first term vanishing as $m \to \infty$.

On the other hand, we have seen if we can separately invert both regimes according to Assumption 2 and equation (3), the structure is

$$\sigma_t = \sum_{k=0}^{m} \bar{\mathcal{P}}(L^0)|u_{t-k-1}|. \tag{17}$$

[8]These coefficients will be higher-order terms of β and γ. Invertibility is ensured by Assumption 2 and the fact that $G_t(\zeta) \in [0,1]$.

By including a sufficient amount of lags of order $m = c\log(n)$ for both models, we can thus test by comparing the model-likelihoods (Vuong, 1989) which specification should be used and if the above extension needs to be applied. We discuss details relating to the implementation of the test in Section 6.

6 Simulation Study

In this section we summarise the results of a comprehensive simulation study. The study is divided into two main parts. First, the proposed asymmetric non-linear smooth transition generalised autoregressive conditional quantile (shortly referred to here and in tables as GACQ instead of ANST-GACQ) procedure will be analysed with respect to different choices of the sample sizes and auxiliary parameters. Later, results are compared with the regime switching GARCH model of Anderson, Nam & Vahid (1999) (shortly labelled as GARCH) for various error distributions, including distributions contaminated by outliers.[9]

By default, the estimation is performed for time series of length $n = 1000$, the number of simulations per experiment is $s = 100$, the composite quantile regression employs by default $k = 9$ quantiles for $\tau \in [0.05, 0.25] \cup [0.75, 0.95]$, the truncation parameter for the ARCH approximation is set to $m = \lceil \frac{5}{2} n^{\frac{1}{4}} \rceil$ and the grid size is $(k_\zeta, k_\eta) = (30, 30)$. The true global parameter vector for both processes is chosen to be $\theta_0 = (\beta_0^{I}, \beta_1^{I}, \gamma_1^{I}, \beta_0^{II}, \beta_1^{II}, \gamma_1^{II})_0^{T} = (0.50, 0.15, 0.60, 0.25, 0.30, 0.15)^{T}$ and the location-scale parameter pair equals $\zeta_0 = (\zeta, \eta)_0^{T} = (0.00, 0.2)^{T}$. While β_0^{I} and β_0^{II} are only determining the unconditional variances of the respective regimes, we chose γ_1^{I} and γ_1^{II} in a way that is consistent with findings in the two regime conditional heteroscedasticity literature (Gonzales-Rivera, 1998; Lubrano, 2001; Wago, 2004; Khemiri, 2011). Unfortunately, the findings on regime-specific parameter values for β_1^{I} and β_1^{II} are rather limited and there is also no clear link to their single regime counterparts. Thus coming up with a sensible prior is somewhat *ad hoc*. We approached this by choosing their values in a way that generates both a higher and a lower persistence regime. Unreported simulations show that different DGPs work similarly well, although, perhaps unsurprisingly, numerical stability deteriorates as one of the regimes' processes becomes close to being integrated.

If not stated otherwise, we will assume the innovations to be standard normally distributed: $\varepsilon_t \sim N(0,1)$. When running simulations using different innovation distributions, in order to ensure comparability, their variances will always be normalised to one. This implies that there is one high and one low variance regime with unconditional variances,

[9] All experiments are conducted using Ox (Doornik, 2009) with extensions written in C for the computationally expensive parts.

defined by $\beta_0^r/(1-\beta_1^r-\gamma_1^r)$ for $r \in \{I, II\}$, of 2 and 0.45, respectively. All of the presented results use a specification with the logistic function G_{logistic}. However, unreported simulations confirmed that the GACQ estimation is insensitive to the misspecification of the transition function (e.g., if the logistic transition function is used while the true underlying model follows the linear or threshold function). Finally, note that we have to restrict the grid for both location and scale. We introduce the data-driven criterion ensuring that location satisfies $\zeta \in [\underline{\zeta}, \bar{\zeta}]$ with unconditional sample quantiles $\underline{\zeta} = \widehat{F}_{u_t}^{-1}(0.1)$ and $\bar{\zeta} = \widehat{F}_{u_t}^{-1}(0.9)$. Similarly, the scale is restricted to $\eta \in [\underline{\eta}, \bar{\eta}(\underline{\zeta}, \bar{\zeta})]$ with fixed $\underline{\eta} = 0.1$ and $\bar{\eta}(\underline{\zeta}, \bar{\zeta}) = \left[\log(0.1^{-1}-1)(0.5\bar{\zeta}-0.5\underline{\zeta})\right]^{-1}$. The latter bound represents the inverse of the logistic function with respect to the scale evaluated at 0.1 and the location at the centre of the considered location grid.

To evaluate the procedures, we report the biases and root mean squared errors (RMSE) of all estimates. As the focus of the quantile regression modelling is on the estimation of quantiles such as Value at Risk rather than parameters, the performance is measured by the mean (absolute) prediction error averaged over the sample, denoted as M(A)PE, absolute one-period-ahead out-of-sample forecast errors (MAFE) and by the coverage ratio, each of them referring to the estimated 5% Value at Risk. Note that the coverage (ratio) is defined as the proportion of observations falling below the estimated Value at Risk and should thus be close to $\tau = 0.05$ for the 5% Value at Risk. It should be mentioned that while coverage, MPE, MAPE and MAFE are reported in the Bias column for the purpose of a tidy exposition, their values represent the mean deviations from the value 0.0, which corresponds to the perfect fit of the model: for example, coverage value 0 represents the exactly correct coverage level 0.05 and MAPE value 0 would represent the exact fit. The RMSE of these quantities additionally depict their corresponding Monte Carlo standard deviations. We will use these metrics to compare different estimators with each other as well as the impact of different features of the data generating process on prediction and forecasting.

Our first simulation experiment, considers the rate of convergence of the proposed estimator by studying its performance for different sample sizes $n = 1000$, $n = 2000$, and $n = 4000$; Table S.5 summarises the results. It is comforting to report that the root mean squared errors (from now on abbreviated as RMSE) of the parameter estimates decrease, at a rate that is consistent with our theoretical conclusions, as the sample size increases. Regarding the second-stage transition parameters, although they are estimated more precisely as the sample size increases too, their RMSEs seem to go down slower than expected. This issue, which will be even more pronounced in the case of the standard GARCH model later, can be caused by the non-linearity of the model with respect to the transition parameters that makes them difficult to estimate from a numerical point of view. This is most pronounced in the first stage, where we use a smooth approximation of the quantile loss function, which is often very flat around the true parameters. Similarly, mean absolute prediction errors (MAPE) and mean absolute forecast errors (MAFE) are decreasing, and unsurprisingly, coverage ratios are accurate by construction of the quantile regression estimator.[10,11]

Moving on to study the influence of auxiliary parameters, to begin with we look at the amount of lags in the ARCH(m) approximation by considering different multiples c of $n^{\frac{1}{4}}$, all satisfying the required order of the ARCH(m) approximation rate. The results for $c = 2.0, 2.5, 3.0$ and 3.5 multiples of $n^{\frac{1}{4}}$, which translate to $m = 12, 15, 17$, and 20 for $n = 1000$, are reported in Table S.6. We conclude they are fairly constant with respect to c, and thus m, although there is a slight U-shape pattern with the optimum in terms of MAPE around $c = 2.5$, which we will use for the remainder of the experiments and the following empirical application.

Further, as discussed in Section 3, the model parameters are not identified at the median and we thus suggested to estimate the first-stage composite quantile regression without using quantiles $\tau \in (0.5 - \delta/2, 0.5 + \delta/2)$ for some $\delta > 0$ as they could introduced extra noise into estimation. In Table S.7, results for different values of δ are collected, indicating that the precision of the estimates seems rather insensitive to a particular choice of δ. For the remainder of the simulations and the empirical application we use $\delta = 0.25$, which corresponds to considering the first and fourth quartile of the data to approximate conditional volatility.

In the second part of the simulation study, we compare smooth transition estimates of conditional quantiles (GACQ) with traditional smooth transition GARCH estimates. In particular, we consider the maximum likelihood estimators of the latter based on both Normal (GARCH-N) and Student's t_4 distribution (GARCH-t).

Naturally, the correctly specified GARCH maximum likelihood estimator yields the best parameter estimates for the case in which the data generating process exhibits standard normally distributed innovations, $\varepsilon_t \sim N(0,1)$; see Table S.8. Neglecting the parameters of the transition function, the GARCH-t model also performs relatively well in terms of RMSEs of the parameter estimates. However, the wrong assumption about the innovation distribution has serious negative consequences on the calculation of conditional quan-

[10]The reason for reporting the coverage ratio is to allow for a direct comparison to the GARCH models in the second part of the simulation study, for which this property does not necessarily hold.

[11]To get an intuition for the magnitude of MAPE and MAFE which are of order 10^{-1}, note that the 5^{th} unconditional quantile of u_t is given by -2.4 for a typical series. The reported statistics refer to integrated absolute deviations of the predictions our model makes for the conditional quantile process which is centred at this value.

tiles and thus its prediction and forecast errors, as can be seen by looking at the GARCH-t estimates in Table S.8. The proposed GACQ model comes with the price of an efficiency loss in the parameter estimates, but the model outperforms both GARCH-N and GARCH-t in terms of predictions errors. While in-sample prediction errors of our model are only slightly smaller than those of GARCH-N and GARCH-t, with respect to out-of-sample forecasting we see a substantial improvement using GACQ over the other two models. This could be partially explained by more precise estimates of the transition function, and in the case of forecasting errors, by directly modelling and fitting the quantiles of the innovation distribution.

The picture is similar for the data generating process with Student errors, $\varepsilon_t \sim t_4/\sqrt{2}$. Again the correctly specified model, in this case GARCH-t, provides the most precise coefficient estimates. Interestingly, the GARCH-N estimator performs better in terms of prediction errors (MAPE) than GARCH-t despite the Student errors. Although the GACQ parameter estimates are less precise than the GARCH ones (with the exception of η), GACQ has similar MAPE as GARCH-N, but outperforms GARCH-N (and thus GARCH-t) in terms of out-of-sample forecasting.

We finalise the innovation-distribution related group of experiments by studying a member of the class of asymmetric distributions. Table S.10 shows results for the case where innovations follow a Gumbel distribution which is parameterised by location parameter $\mu_G = 0$ and scale parameter $\beta_G = \sqrt{6}/\pi$. We re-centred the innovations by subtracting $\beta_G e^1$ from each realisation so that ε has mean zero. Since it is distribution-agnostic, it should come with no surprise that the performance of the proposed GACQ model is similar to the previous experiments with symmetric errors. Being misspecified, the GARCH-N and GARCH-t models provide now less precise estimates in regime II and both their MAPEs and MAFEs are larger than those of GACQ.

Finally, we look at the case in which normally distributed innovations are contaminated by outliers. We define them as follows. Let $\varepsilon_t \sim N(0,1)$ or t_4 and $r_t \sim \mathcal{U}[0,1]$ are independent and uniformly distributed. Then for each $u_t(\theta_0) = \sigma_t(z_t, \theta_0)\varepsilon_t$, the contaminated series $\{u'\}_{t=1}^n$ is defined as

$$u'_t := u_t + \mathbb{1}_{\{r_t \leqslant 0.025\}} \mathrm{sgn}(\varepsilon_t) 3\sigma_\varepsilon$$

with $\sigma_\varepsilon = 1$. Note that this might be considered a very small contamination, but we report estimates for the 5% Value at Risk and thus these contaminated values form a large proportion of the data used for estimation. Predictably, the RMSEs of all the parameters estimates increase in the presence of contaminations irrespective of the considered model. Considering MAPE, the increase in the prediction errors is relatively limited in the case of the GACQ method. On the other hand, the MAPE of GARCH-N increase substantially and the model exhibits large prediction biases. The situation is similar for GARCH-t, which however partially compensates for this fatter tails by underestimating its degrees of freedom parameter. Most importantly, the coverage ratio in the conditional variance GARCH models are on average off by 1.4 and 2.7 percentage points, respectively, which is a rather significant deviation given that we consider the 5% Value at Risk, whereas the coverage ration of GACQ is unaffected by the contamination.

7 Empirical Application

For this empirical study we consider the 1% and 5% Value at Risk of daily closing data of the USD/GBP exchange rate and the German equity index (DAX).

Before we estimate the GACQ model with the smooth transition specification (3), we first test this specification against the alternative of the Anderson, Nam & Vahid (1999) specification (8). For the implementation of this test we specify a quadratic GARCH model as defined in Section 2. Under the null hypothesis, both regimes can be inverted separately and approximated by an ARCH(m) process according to equation (17). Otherwise, the lag polynomials will be dependent on higher-order interactions of the transition function as derived in Lemma (1) leading to a different ARCH(m) specification. Table 1 presents the exact specifications of the estimated lag polynomials of $G_t(\zeta)$ and $1 - G_t(\zeta)$ under the null and the alternative hypothesis, respectively, as well as likelihood ratio test statistics which we obtain from maximum likelihood estimation of both specifications. We do not find evidence in favour of the non–separable specification for both time series and thus proceed with our baseline estimation procedure as discussed in Section 3.

Table 2 reports GACQ parameter estimates for the two series using data from 2002 to 2016 and a corresponding sample size of $n = 3{,}000$. Model selection was based on a formal selection criterion based on the loss resulting from the second step of the estimation procedure. The specification that is ultimately reported corresponds to the widely used specification with $p = q = 1$ lags and the logistic transition function, although estimates turned out to be very robust with respect to the choice of the latter. Results are reported for the transition variable $\xi_t = u_{t-1}$.[12]

Recall that these parameters are local with respect to the estimated innovation distri-

[12]Alternative models in ascending order or their loss for the respective 1% VaR are: 38.056 (u_{t-1}), 39.265 (u_{t-2}), 40.512 ($\bar{u}_t^5$) and 40.515 (u_{t-3}) for USD/GBP and 129.41 (u_{t-3}), 133.15 (u_{t-1}), 133.42 ($\bar{u}_t^5$), 134.58 (u_{t-2}) for the DAX where $\bar{u}_t^5 := \sum_{k=1}^{5} u_{t-k}$ are weekly average returns

Table 1: Polynomials, Likelihoods and Critical values of Specification Test

Lags	H_0	H_1
$1, u_{t-1}$	G_t, G_t^-	G_t, G_t^-
u_{t-2}	G_t, G_t^-	$G_tG_{t-1}, G_tG_{t-1}^-, G_t^-G_{t-1}, G_t^-G_{t-1}^-$
u_{t-3}	G_t, G_t^-	$G_tG_{t-1}G_{t-2}, G_tG_{t-1}G_{t-2}^-, G_tG_{t-1}^-G_{t-2}, G_tG_{t-1}^-G_{t-2}^-$
		$G_t^-G_{t-1}G_{t-2}, G_t^-G_{t-1}G_{t-2}^-, G_t^-G_{t-1}^-G_{t-2}, G_t^-G_{t-1}^-G_{t-2}^-$
u_{t-4}	G_t, G_t^-	$G_tG_{t-1}G_{t-2}G_{t-3}, G_tG_{t-1}G_{t-2}G_{t-3}^-, G_tG_{t-1}G_{t-2}^-G_{t-3}, G_tG_{t-1}G_{t-2}^-G_{t-3}^-$
		$G_tG_{t-1}^-G_{t-2}G_{t-3}, G_tG_{t-1}^-G_{t-2}G_{t-3}^-, G_tG_{t-1}^-G_{t-2}^-G_{t-3}, G_tG_{t-1}^-G_{t-2}^-G_{t-3}^-$
		$G_t^-G_{t-1}G_{t-2}G_{t-3}, G_t^-G_{t-1}G_{t-2}G_{t-3}^-, G_t^-G_{t-1}G_{t-2}^-G_{t-3}, G_t^-G_{t-1}G_{t-2}^-G_{t-3}^-$
		$G_t^-G_{t-1}^-G_{t-2}G_{t-3}, G_t^-G_{t-1}^-G_{t-2}G_{t-3}^-, G_t^-G_{t-1}^-G_{t-2}^-G_{t-3}, G_t^-G_{t-1}^-G_{t-2}^-G_{t-3}^-$
USDGBP	-1360.4	-1347.8 ($\chi^2_{0.95,22} = 33.9 > 25.2 = \text{LR}$)
DAX	11606.5	11353.9 (H_0 model superior to H_1 (*))

Note: This table shows the included interaction terms according to Lemma (1), where we denote $G_t = G_t(\zeta)$ and $G_t^- = 1 - G_t(\zeta)$, according to the test derived in Section 5 for both the specification under the null hypothesis (separability) and the alternative. It reports the respective log-likelihoods for both time series and the results of a LR test. Tests were repeated for $m = 3$, with the same conclusion.
(*) Note that H_0 is not nested in H_1. LR tests of non–nested hypotheses are studied in Vuong (1989).

bution, which results in negative estimates.[13] Given that we consider financial time series, which are known to react to news relatively quickly, it is not surprising that in most cases the regime is determined by the previous observation of the lagged dependent variable. This is with the exception of the 1% Value at Risk model for the DAX, for which our model selection procedure suggests that the regime is determined by u_{t-3}.

With -0.08% and -0.10%, corresponding to the 36^{th} and 33^{rd} unconditional quantile, respectively, the location parameters for the USD/GBP are rather close to zero. The scale parameter estimates suggest that there is a rapid transition from one regime to another around these locations.[14] The first regime is active more often, namely for about 62% of the daily observations.[15] The unconditional variance of the first regime is 0.19 and thus approximately the same as the one for the second regime for which we obtain an estimate of 0.14. Based on the ARCH parameter estimates from the first stage, we do not find one regime to be more persistent than the other.

The estimated locations for the German equity index are 0.72% for the 5% VaR and -1.09% for the 1% VaR, corresponding to the 74^{th} and 15^{nd} unconditional quantile, respectively. The scale estimate also implies a rapid transition from one regime to the other around the location. Finally the first regime of the DAX is active 40% of the time with an unconditional variance of 0.45. The variance of the second regime is higher with 0.59. Again,

[13]Note that one could theoretically apply another composite quantile regression procedure for the second stage to identify the global GARCH parameters.
[14]No significance levels of the location and scale parameter estimates are reported, due to the lack of a natural null hypothesis for either of them.
[15]We define this measure as an unconditional average of the value of the transition function.

Table 2: Coefficients and standard errors for the Value at Risk GACQ estimates.

1% Value at Risk

	USD/GBP ($\xi = u_{t-1}$)					DAX ($\xi = u_{t-1}$)				
	Regime I		Regime II			Regime I		Regime II		
	coef	s.e.	coef	s.e.	$P(>\|Z\|)$	coef	s.e.	coef	s.e.	$P(>\|Z\|)$
β_0	-0.021	0.087	-0.556	0.365	0.012	-0.029	0.497	-0.771	0.123	0.003
β_1	-1.557	0.371	-0.000	0.405	0.000	-1.846	0.225	-1.108	0.163	0.000
γ_1	-0.011	0.165	-0.918	0.558	0.038	-0.000	0.305	-0.237	0.084	0.009
ζ	-0.081	0.383				0.725	0.523			
η	0.100	0.268				0.100	0.355			

5% Value at Risk

	USD/GBP ($\xi = u_{t-1}$)					DAX ($\xi = u_{t-1}$)				
	Regime I		Regime II			Regime I		Regime II		
	coef	s.e.	coef	s.e.	$P(>\|Z\|)$	coef	s.e.	coef	s.e.	$P(>\|Z\|)$
β_0	-0.242	0.057	-0.783	0.125	0.000	-0.809	0.162	-2.832	0.780	0.000
β_1	-2.084	0.168	-0.000	0.216	0.000	-2.367	0.200	-0.093	0.450	0.000
γ_1	-0.000	0.064	-1.240	0.129	0.000	-0.000	0.133	-0.283	0.128	0.000
ζ	-0.108	0.091				-1.096	0.670			
η	0.100	0.078				0.131	0.564			

Note: The parameters $\beta_0^\tau, \beta_1^\tau, \gamma_1^\tau$ for $r \in \{I, II\}$ are local with respect to τ and estimated subject to non-positivity constraints (Koenker & Ng, 2005). The reported p-values correspond to the test of the hypothesis that individual parameters are different in the two respective regimes. Note that, technically these asymptotic tests are only valid for the parameters in the interior of the parameter space, i.e. whenever the linear inequality constraint is not binding.

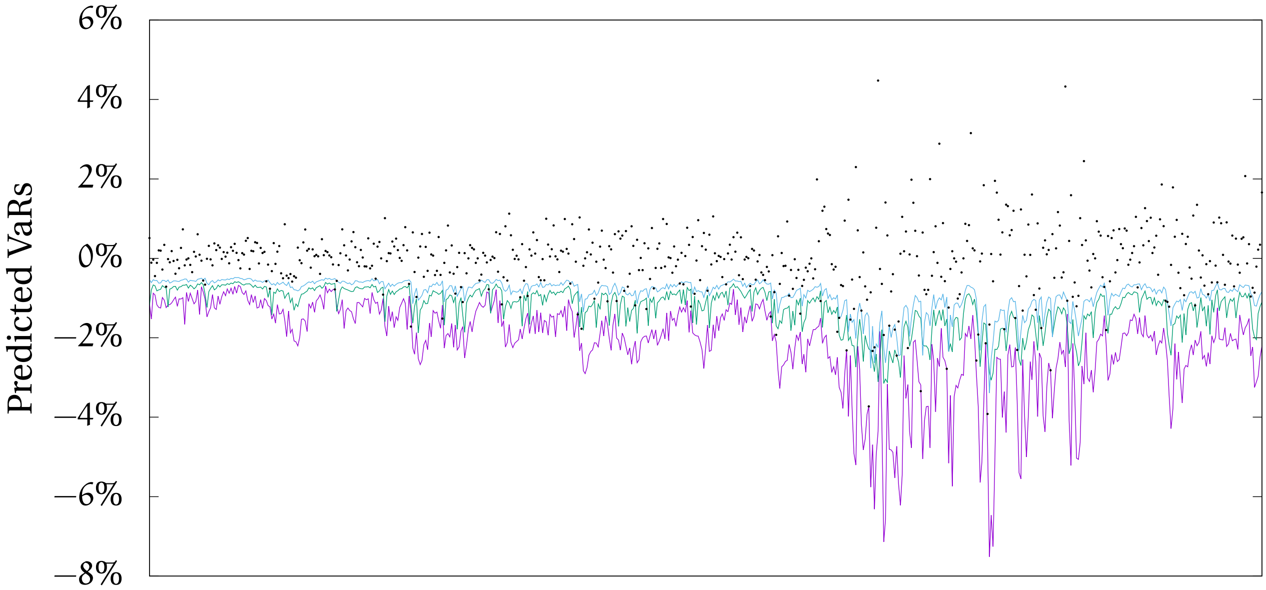
there is no clear order of persistence between the regimes.

To conclude this first part of our application, Figure 1 presents the daily returns of exchange rate pair USD/GBP as well as the corresponding predicted 10%, 5%, and 1% value at risk processes according to our estimation procedure.[16] The main conclusions we can draw from this are that the processes are indeed stationary, and compared to GARCH-based models, they are not proportional to each other due to the non-linear nature of the model in combination with the fact that estimates are local with respect to the considered quantile of the innovations.

To compare the performance of the GACQ Value at Risk estimates, GARCH estimates, and one regime GACQ estimates, we will now apply two different out of sample tests which are based on the coverage ratios. For this, we consider a rolling sub-sample of both time series, where we evaluate the forecast for each t within the last $N = 100$ periods. For each

[16] Technically the Basel regulation uses the actuarial convention that value at risk has a positive sign when it corresponds to a loss. According to this definition Figure 1 shows estimates of "negative VaRs".

Figure 1: USD/GBP returns and its (ANST-GACQ) predicted 10%, 5% and 1% VaR's



forecast we use the previous 1000 observations to obtain model estimates, based on which we construct forecasts for the conditional quantile $\widehat{Q}_{u_t}(\tau|\mathcal{F}_{t-1})$ for $\tau \in (0,1)$. Using the true value u_t we calculate $I_t = \mathbb{1}\left\{u_t \leqslant \widehat{Q}_{u_t}(\tau|\mathcal{F}_{t-1})\right\}$ for each $t = n - N + 1, \ldots, n$, where n denotes the latest observation in the data. The out-of-sample coverage ratio is then equal to $h_{oos} := \sum_{t=n-N+1}^{n} I_t/N$. We use two different tests to verify that the proportion of forecasts exceeding the estimated quantile, also referred to as "hits," is not significantly different from the specified τ. Both tests are widely used in the literature and evaluate unconditional coverage.[17] The first test we use is the likelihood ratio test proposed by Kupiec (1995). It assumes that the Bernoulli process I_t is an i.i.d. sequence, so that the number of hits $x = Nh_{oos}$ follows a binomial distribution with $h_{oos} = \tau$ under the null hypothesis. The second test to verify that the out-of-sample forecasts represent the τ^{th} quantile exploits the fact that $I_{t+1} - \tau$ is a martingale difference sequence with zero mean and variance $N\tau(1-\tau)$ and thus its cumulative sum converges to a normal distribution (Campbell, 2007; Xiao & Koenker, 2009). Detailed information and formal definitions of the tests statistics can be found in Appendix **??**. Table 3 summarizes the results for the two regime models.

It is apparent that the GACQ model estimates conditional quantiles more accurately than both GARCH models and there is no evidence that the null hypothesis, stating that the forecast represents the true quantile, has to be rejected for any of the confidence levels. This is not the case for the regime-switching quadratic GARCH-N and GARCH-t models. While there are situations in which these models also perform well, what we would take away from this experiment is that our GACQ estimator should be preferred with regards to

[17] There are other tests which also consider conditional coverage, discussed for example in Christoffersen (1998), Berkowitz, Christoffersen & Pelletier (2011) or the dynamic quantile test proposed in Engle & Manganelli (2004).

Table 3: Coverage and test statistics for GACQ and GARCH with two regimes.

USD/GBP Exchange rate

	GACQ			GARCH-N			GARCH-t		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
h_{IS}	0.012	0.052	0.101	0.016	0.046	0.089	0.001	0.012	0.028
h_{OOS}	0.000	0.030	0.070	0.000	0.030	0.060	0.000	0.010	0.040
$P(>\|Z_n\|)$	0.315	0.359	0.317	0.315	0.359	0.183	0.315	0.067	0.046
$P(>\mathcal{L}_n)$	0.156	0.323	0.293	0.156	0.323	0.153	0.156	0.026	0.024

DAX Equity Index

	GACQ			GARCH-N			GARCH-t		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
h_{IS}	0.011	0.050	0.010	0.030	0.089	0.121	0.012	0.138	0.171
h_{OOS}	0.000	0.060	0.100	0.040	0.110	0.150	0.020	0.080	0.150
$P(>\|Z_n\|)$	0.315	0.646	1.000	0.003	0.006	0.096	0.315	0.169	0.096
$P(>\mathcal{L}_n)$	0.156	0.656	1.000	0.023	0.017	0.118	0.376	0.204	0.118

Table 4: Coverage and test statistics for GACQ with one regime.

	USD/GBP				DAX Equity Index			
Quantile	h_{IS}	h_{OOS}	$P(>\|Z_n\|)$	$P(>\mathcal{L}_n)$	h_{IS}	h_{OOS}	$P(>\|Z_n\|)$	$P(>\mathcal{L}_n)$
1%	0.012	0.000	1.000	1.000	0.011	0.010	1.000	1.000
5%	0.050	0.010	0.067	0.026	0.050	0.060	0.646	0.656
10%	0.099	0.050	0.096	0.068	0.099	0.140	0.183	0.206

its performance uniformly over different types of time series, as well as different estimated quantiles for each of these series.

One could of course argue that the two regime models such as GACQ and GARCH are unnecessary and a single regime CAViaR model would be sufficient. Therefore, we also compare the two regime GACQ model with the single regime GACQ (CAViaR) by means of analysing their out-of-sample forecasts. Results for the single regime model are reported in Table 4 and are to be contrasted with the first three columns of Table 3. While it would in principal be possible to develop a formal test for the regime-switching model against the linear model (e.g., see Luukkonen, Saikkonen & Teraesvirta (1988) for the smooth transition autoregressive model), we think this exceeds the scope of this paper.

The results of this final part of our experiments again suggest different conclusions for each of the two time series. For USD/GBP, the hypothesis that the proportion of hits equals the specified Value at Risk in the out-of-sample forecast procedure can be rejected for the

5% and 10% quantiles. This is not the case for the two regime GACQ model, which is a strong indication that there is indeed a second regime in this time-series. On the other hand we cannot reject the hypothesis that the one-regime GACQ model does not describe the German equity index equally well as its two regime counterpart.

8 Conclusion

In the preceding analysis we proposed a model which bridges the gap between recent developments in Value at Risk estimation such as CAViaR models and traditional smooth transition GARCH models. We believe some general conclusions can be drawn from our experiments. First, from the case of USD/GBP we see that ignoring a second regime in situations where the DGP seems to include one, unsurprisingly leads to inferior out of sample predictions. Second, from the results for the DAX equity index we conclude that while we find the parameter estimates of both regimes to be different suggesting the existence of two regimes, specifying only one works equally well in terms of out-of-sample forecast errors for this particular time series. Despite this and in combination with the fact that we needed a technical identification assumption of the existence of two regimes for our theoretical results it is comforting to see that this does not harm the forecasting accuracy of our estimator. Third, confirming the conclusions already found in some of the related literature (see Xiao & Koenker (2009) and references therein), compared to their conditional variance counterparts, a distribution-free specification and direct estimation of conditional quantiles seems to work well in practice. From all these points we conclude that while other models and specifications do work for data generating processes unknown to the empirical analyst, estimating the value at risk using the proposed GACQ model reduces the risk of misspecification, since it works equally well and uniformly over different DGPs.

References

Giovanni Barone ADESI, Kostas GIANNOPOULOS & Les VOSPER (1999), VaR without correlations for portfolios of derivative securities. *Journal of Future Markets*, vol. 19, no. 5, pp. 583–602.

C.O ALEXANDER & C.T LEIGH (1997), On the Covariance Matrices Used in Value at Risk Models. *The Journal of Derivatives*, vol. 4, no. 3, pp. 50–62.

Torben ANDERSEN & Tim BOLLERSLEV (2006), Volatility and correlation forecasting. *Journal of Economic Forecasting*, vol. 1, pp. 777–878.

Heather M. ANDERSON, Kiseok NAM & Farshid VAHID (1999), Asymmetric nonlinear smooth transition GARCH models. *Dynamic Modeling and Econometrics in Economics and Finance*, vol. 1, pp. 191–207.

Jeremy BERKOWITZ, Peter CHRISTOFFERSEN & Denis PELLETIER (2011), Evaluating Value-at-Risk Models with Desk-Level Data. *Management Science*, vol. 57, no. 12, pp. 2213–2227.

Fischer BLACK (1976), Studies of stock price volatility changes. *Proceedings of the 1976 Meetings of the American Statistical Association*, pp. 177–181.

Tim BOLLERSLEV (1986), Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, vol. 31, pp. 307–327.

Zongwu CAI & Xiaoping XU (2009), Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association*, vol. 104, no. 485, pp. 371–383.

Sean CAMPBELL (2007), A review of backtesting and backtesting procedures. *The Journal of Risk*, vol. 9, no. 2, pp. 1–17.

Marine CARRASCO & Xiaohong CHEN (2002), Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, vol. 18, no. 1, pp. 17–39.

Felix CHAN & Michael MCALEER (2003), Estimating smooth transition autoregressive models with GARCH errors in the presence of extreme observations and outliers. *Applied Financial Economics*, vol. 13, no. 8, pp. 581–592.

Xiaohong CHEN (2008), Large Sample Sieve Estimation of Semi-Nonparametric Models. In Handbook of Econometrics (edited by James J. HECKMAN & Edward E. LEAMER), volume 6b ed., Elsevier, ISBN 978-0-444-53200-8, pp. 5549–5633.

Xiaohong CHEN & Xiaotong SHEN (1998), Sieve Extremum Estimates for Weakly Dependent Data. *Econometrica*, vol. 66, no. 2, pp. 289–314.

Peter F. CHRISTOFFERSEN (1998), Evaluating Interval Forecasts. *International Economic Review*, vol. 39, no. 4, p. 841.

Jurgen DOORNIK (2009), An Object-Oriented Matrix Programming Language Ox 6. London: Timberlake Consultants Ltd.

Robert F. ENGLE (1982), Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, vol. 50, no. 4, pp. 987–1007.

Robert F ENGLE & Simone MANGANELLI (2004), CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business and Economic Statistics*, vol. 22, no. 4, pp. 367–381.

R FREY & A MCNEIL (1998), Estimation of Tail-Related Risk Measures for Heteroskedastic Financial Time Series: An Extreme Value Approach. *Journal of Empirical Finance*, vol. 7, pp. 271–300.

Richard H. GERLACH, Cathy W. S. CHEN & Nancy Y. C. CHAN (2011), Bayesian Time-Varying Quantile Forecasting for Value-at-Risk in Financial Markets. *Journal of Business & Economic Statistics*, vol. 29, no. 4, pp. 481–492.

Lawrence R. GLOSTEN, Ravi JAGANNATHAN & David E. RUNKLE (1993), On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, vol. 48, no. 5, pp. 1779–1801.

Gloria GONZALES-RIVERA (1998), Smooth-Transition GARCH Models. *Studies in Nonlinear Dynamics and Econometrics*, vol. 3, no. 2, pp. 61–78.

C. GOURIEROUX, J. P. LAURENT & O. SCAILLET (2000), Sensitivity analysis of Values at Risk. *Journal of Empirical Finance*, vol. 7, no. 3-4, pp. 225–245.

Dashan HUANG, Baimin YU, Frank J. FABOZZI & Masao FUKUSHIMA (2009), CAViaR-based forecast for oil price risk. *Energy Economics*, vol. 31, no. 4, pp. 511–518.

Peter J. HUBER (1964), Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101.

Abdulobič IBRAGIMOV & Yuri LINNIK (1971), Independent and stationary sequences of random variables.

Rim KHEMIRI (2011), The smooth transition GARCH model: application to international stock indexes. *Applied Financial Economics*, vol. 21, no. 8, pp. 555–562.

Franc KLAASSEN (2002), Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics*, vol. 27, no. 2, pp. 363–394.

Roger KOENKER (2005), Quantile regression. 1 ed., Cambridge: Econometric Society Monographs, ISBN 13-978-0-521-84573.

Roger KOENKER & Gilbert BASSETT (1978), Regression quantiles. *Econometrica*, vol. 46, no. 1, pp. 33–50.

Roger KOENKER & Pin NG (2005), Inequality constrained quantile regression. *Sankhya: The Indian Journal of Statistics*, vol. 67, no. 2, pp. 418–440.

Roger KOENKER & Quanshui ZHAO (1996), Conditional quantile estimation and inference for ARCH models. *Econometric Theory*, vol. 12, no. 5, pp. 793–813.

Paul KUPIEC (1995), Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, vol. 3, no. 2, pp. 73–84.

Sangyeol LEE & Jungsik NOH (2013), Quantile Regression Estimator for GARCH Models. *Scandinavian Journal of Statistics*, vol. 40, no. 1, pp. 2–20.

Wai Keung LI & C. W. LI (1996), On a double-threshold autoregressive heteroscedastic time series model. *Journal of Applied Econometrics*, vol. 11, pp. 253–274.

Michel LUBRANO (2001), Smooth transition GARCH models: A Bayesian perspective. *Louvain Economic Review*, vol. 67, no. 3, pp. 257–288.

Ritva LUUKKONEN, Pentti SAIKKONEN & Timo TERAESVIRTA (1988), Testing linearity against smooth transition autoregressive models. *Biometrika*, vol. 75, no. 3, pp. 491–499.

Grant MCQUEEN & Keith VORKINK (2004), Whence GARCH? A Preference-Based Explanation for Conditional Volatility. *The Review of Financial Studies*, vol. 17, no. 4, pp. 915–949.

Mika MEITZ & Pentti SAIKKONEN (2008), Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models. *Econometric Theory*, vol. 24, pp. 1291–1320.

Antonio MELE & Fabio FORNARI (1997), Sign - and volatility - switching arch models: theory and applications to international stock markets. *Journal of applied econometrics*, vol. 12, no. 1, pp. 49–65.

Kiseok NAM, Chong Soo PYUN & Stephen L. AVARD (2001), Asymmetric reverting behavior of short-horizon stock returns: An evidence of stock market overreaction. *Journal of Banking & Finance*, vol. 25, no. 4, pp. 807–824.

Whitney K NEWEY & Daniel MCFADDEN (1994), Large Sample Estimation and Hypothesis Testing. *Handbook of Econometrics*, vol. 4, no. 36, pp. 2112–2245.

R. RABEMANANJARA & J. M. ZAKOIAN (1993), Threshold arch models and asymmetries in volatility. *Journal of Applied Econometrics*, vol. 8, no. 1, pp. 31–49.

Matthew P. RICHARDSON, Jacob BOUDOUKH & Robert WHITELAW (1998), The Best of Both Worlds: A Hybrid Approach to Calculating Value at Risk. *SSRN*, vol. Jan, no. 51420, pp. 1–12.

O. SCAILLET (2004), Nonparametric estimation and sensitivity analysis of expected shortfall. *Mathematical Finance*, vol. 14, no. 1, pp. 115–129.

G. William SCHWERT (1990), Stock Volatility and the Crash of'87. *Review of Financial Studies*, vol. 3, no. 1, pp. 77–102.

Steven TAYLOR (1986), Modelling Financial Time Series. Chichester: Wiley.

Timo TERASVIRTA (1992), Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics*, vol. 7, pp. 119–136.

Howell TONG & K.S. LIM (1980), Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 42, pp. 245–292.

Aad VAN DER VAART & Jon WELLNER (1996), Weak Convergence and Empirical Processes: With Applications to Statistics. New York: Springer.

Quang H. VUONG (1989), Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, vol. 57, no. 2, pp. 307–333.

Hajime WAGO (2004), Bayesian estimation of smooth transition GARCH model using Gibbs sampling. *Mathematics and Computers in Simulation*, vol. 64, no. 1, pp. 63–78.

Halbert WHITE, Kim TAE-HWAN & Simone MANGANELLI (2008), Modeling autoregressive conditional skewness and kurtosis with multi-quantile CAViaR. *ECB Working Paper*, vol. 957.

Zhijie XIAO, Hongtao GUO & Miranda S. LAM (2015), Handbook of Financial Econometrics and Statistics. New York, NY: Springer New York, ISBN 978-1-4614-7749-5.

Zhijie XIAO & Roger KOENKER (2009), Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models. *Journal of the American Statistical Association*, vol. 104, no. 488, pp. 1696–1712.

Zhibiao ZHAO & Zhijie XIAO (2014), Efficient Regressions via Optimally Combining Quantile Information. *Econometric theory*, vol. 30, no. 6, pp. 1272–1314.

Songfeng ZHENG (2011), Gradient descent algorithms for quantile regression with smooth approximation. *International Journal of Machine Learning and Cybernetics*, vol. 2, no. 3, pp. 191–207.

Hui ZOU & Ming YUAN (2008), Composite quantile regression and the oracle model selection theory. *The Annals of Statistics*, vol. 36, no. 3, pp. 1108–1126.

A.1 Auxiliary Results and Proofs

Within the following derivations, let $\mathbb{P}_n$ be the empirical distribution that puts mass $d\mathbb{P}_n = n^{-1}$ to each observation $u_1, \dots, u_n$ such that $\mathbb{P}_n f(u_t) = \int f(u_t) d\mathbb{P}_n = n^{-1} \sum_{t=1}^{n} f(u_t)$ for any measurable function f. Also let the vector $z_t^m = (|u_{t-1}|, \dots, |u_{t-m}|)^T$, and $z_t^m(\zeta) = [G_t(\zeta)\, z_t^m, (1 - G_t(\zeta)) z_t^m]^T$, where the latter is defined as a function of ζ to highlight the dependence on the transition function. Further, let the data considered in the second stage be $z_t(\alpha) = [1, |u_{t-1}|, \dots, |u_{t-p}|, \sigma_{t-1}(\alpha), \dots, \sigma_{t-q}(\alpha)]^T$ which is a function of $\alpha := [\alpha^{I,T}, \alpha^{II,T}, \zeta^T]^T$, where ζ refers to the location and scale parameters entering the first stage. Similarly, we define $z_t(\alpha, \zeta) = [G_t(\zeta)\, z_t(\alpha), (1 - G_t(\zeta)) z_t(\alpha)]^T$ to be the vector that stacks both regimes weighted data. Also recall that $a = [\alpha^{I,T}, \alpha^{II,T}, q^T, \zeta^T]^T$.

Now the right-side derivative of the check-function $\rho_{\tau_k}(u) = u(\tau_k - \mathbb{1}\{u \leqslant 0\})$ is given by $\psi_{\tau_k,t}(u) := (\tau_k - \mathbb{1}\{u \leqslant 0\})$ so that the directional derivative of the objective function defined in equation (9) is given as

$$g_n(a) = g_n(\alpha^I, \alpha^{II}, q, \zeta) := \mathbb{P}_n \sum_{k=1}^{K} \underbrace{\begin{bmatrix} z_t^m(\zeta) q_k \\ \mathbb{1}\{k = 1\}[\alpha^I, \alpha^{II}]^T z_t^m(\zeta) \\ \vdots \\ \mathbb{1}\{k = K\}[\alpha^I, \alpha^{II}]^T z_t^m(\zeta) \\ q_k(\alpha^I - \alpha^{II})^T z_t^m(\zeta) \frac{\partial G_t(\zeta)}{\partial \zeta} \end{bmatrix}}_{x_{t,k}(a)} (\tau_k - \mathbb{1}\{u_t \leqslant q_k[\alpha^I, \alpha^{II}]^T z_t^m(\zeta)\}). \tag{18}$$

Similarly, without making any statements about convergence yet, the corresponding population equivalent can be written as

$$g(a) = g(\alpha^I, \alpha^{II}, q, \zeta) := \mathbb{E} \sum_{k=1}^{K} \begin{bmatrix} z_t^m(\zeta) q_k \\ \mathbb{1}\{k = 1\}[\alpha^I, \alpha^{II}]^T z_t^m(\zeta) \\ \vdots \\ \mathbb{1}\{k = K\}[\alpha^I, \alpha^{II}]^T z_t^m(\zeta) \\ q_k(\alpha^I - \alpha^{II})^T z_t^m(\zeta) \frac{\partial G_t(\zeta)}{\partial \zeta} \end{bmatrix} (\tau_k - F_{u_t|\mathcal{F}_{t-1}}(q_k[\alpha^I, \alpha^{II}]^T z_t^m(\zeta)))$$

by applying the law of iterated expectations. In addition to this we will frequently make use of the identity $\partial F_{u_t|\mathcal{F}_{t-1}}(x)/\partial x = \sigma_t{}^{-1} \partial F_\varepsilon(x)/\partial x = \sigma_t^{-1} f_\varepsilon(x)$: as equation (4) establishes $F^{-1}_{u_t|\mathcal{F}_{t-1}}(x) = \sigma_t F_\varepsilon^{-1}(x)$ and since both $F_{u_t|\mathcal{F}_{t-1}}$ and F_ε are monotone and differentiable by Assumption 3 and 4, the expression follows by applying the inverse function theorem.

Proof of Lemma 1: We show this by induction and omit the parameters of sigma for readability. By definition we have

$$\sigma_t = \bar{\mathcal{P}}(L^0)\sigma_{t-1} + \bar{\mathcal{P}}(L^0)|u_{t-1}|.$$

Lagging this equation and plugging in for σ_{t-1}, we get by construction of $\bar{\mathcal{P}}(L^1)$:

$$\begin{aligned}\sigma_t &= \bar{\mathcal{P}}(L^0)\left[L\bar{\mathcal{P}}(L^0)\sigma_{t-2} + L\bar{\mathcal{P}}(L^0)|u_{t-2}|\right] + \bar{\mathcal{P}}(L^0)|u_{t-1}| \\ &= \bar{\mathcal{P}}(L^1)\sigma_{t-2} + \bar{\mathcal{P}}(L^1)|u_{t-2}| + \bar{\mathcal{P}}(L^0)|u_{t-1}|.\end{aligned}$$

For $m \mapsto m+1$ we note that $\sigma_{t-j} = L^j\bar{\mathcal{P}}(L^0)\sigma_{t-j-1} + L^j\bar{\mathcal{P}}(L^0)|u_{t-j-1}|$. Thus substituting for σ_{t-m-1} in equation (16) we obtain

$$\sigma_t = \bar{\mathcal{P}}(L^m)\left[L^{m+1}\bar{\mathcal{P}}(L^0)\sigma_{t-m} + L^{m+1}\bar{\mathcal{P}}(L^0)|u_{t-m}|\right] + \sum_{k=0}^{m}\bar{\mathcal{P}}(L^k)|u_{t-k-1}|,$$

and after using $L^{m+1}\bar{\mathcal{P}}(L^0) = L\bar{\mathcal{P}}(L^m)$ we get

$$\begin{aligned}\sigma_t &= \bar{\mathcal{P}}(L^{m+1})\sigma_{t-(m+1)-1} + \bar{\mathcal{P}}(L^{m+1})|u_{t-(m+1)-1}| + \sum_{k=0}^{m}\bar{\mathcal{P}}(L^k)|u_{t-k-1}| \\ &= \bar{\mathcal{P}}(L^{m+1})\sigma_{t-(m+1)-1} + \sum_{k=0}^{m+1}\bar{\mathcal{P}}(L^k)|u_{t-k-1}|,\end{aligned}$$

where the first term is vanishing due to the parameter restrictions in Assumption 2 and $G_t(\zeta) \in [0,1]$ by Assumption 1, which completes the proof. □

Proof of Theorem 1: The proof is split into three parts: we first discuss the identification of the first-stage parameters, then their consistency, and finally, the convergence rate is derived. First, we show that for $\boldsymbol{\alpha}(\tau) = [\boldsymbol{\alpha}^{I,T}(\tau), \boldsymbol{\alpha}^{II,T}(\tau)]$,[18] the vector $\left[\boldsymbol{\alpha}_0^T(\tau), \zeta_0^T\right]^T$ is in fact the minimum of the objective function $\mathbb{E}\rho_\tau(u)$. We show this for an arbitrary quantile $\tau \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1)$, from which the composite quantile result follows as $K > 1$ by Assumption 7.

To see this let the objective function $m(\boldsymbol{\alpha}(\tau), \zeta) = \mathbb{E}\rho_\tau(u_t - \boldsymbol{\alpha}^T(\tau)\boldsymbol{z}_t^m(\zeta))$. Then we have global identification if and only if $m(\boldsymbol{\alpha}(\tau), \zeta) - m(\boldsymbol{\alpha}_0(\tau), \zeta_0) > 0$ for any $\left[\boldsymbol{\alpha}^T(\tau), \zeta^T\right] \neq \left[\boldsymbol{\alpha}_0^T(\tau), \zeta_0^T\right]$.

Let $\boldsymbol{\alpha}(\tau)^\Delta = \boldsymbol{\alpha}^I(\tau) - \boldsymbol{\alpha}^{II}(\tau)$ and also let ν_t be the probability measure of u_t conditional upon the filtration $\mathcal{F}_{t-1}$. Then we have to prove

$$\inf_{\substack{\boldsymbol{\alpha}(\tau)\,:\,\|\boldsymbol{\alpha}(\tau)-\boldsymbol{\alpha}_0(\tau)\|>\delta \\ \zeta\,:\,\|\zeta-\zeta_0\|>\delta}} \mathbb{E}\left[\int \rho_\tau(u_t - \boldsymbol{\alpha}(\tau)^T\boldsymbol{z}_t^m(\zeta))d\nu_t - \int \rho_\tau(u_t - \boldsymbol{\alpha}_0(\tau)^T\boldsymbol{z}_t^m(\zeta_0))d\nu_t\right] > \varepsilon_\delta, \tag{19}$$

where we can discuss the inside part of the expectation into two cases. First, consider $\boldsymbol{\alpha}_0(\tau)^T\boldsymbol{z}_t^m(\zeta_0) >$

[18] In the first stage, the parameters of the transition function are estimated globally using the objective function at quantiles $\tau_1, \ldots, \tau_K$ and are therefore not included in the locally estimated $\boldsymbol{\alpha}(\tau)$.

$\alpha(\tau)^\top z_t^m(\zeta)$:

$$
\begin{aligned}
&\int \rho_\tau(u_t - \alpha(\tau)^\top z_t^m(\zeta))d\nu_t - \int \rho_\tau(u_t - \alpha_0(\tau)^\top z_t^m(\zeta_0))d\nu_t = \\
&(\tau-1)\int_{-\infty}^{\alpha(\tau)^\top z_t^m(\zeta)}(u_t - \alpha(\tau)^\top z_t^m(\zeta))d\nu_t + \tau\int_{\alpha(\tau)^\top z_t^m(\zeta)}^{\alpha_0(\tau)^\top z_t^m(\zeta_0)}(u_t - \alpha(\tau)^\top z_t^m(\zeta))d\nu_t + \tau\int_{\alpha_0(\tau)^\top z_t^m(\zeta_0)}^{+\infty}(u_t - \alpha(\tau)^\top z_t^m(\zeta))d\nu_t \\
&-(\tau-1)\int_{-\infty}^{\alpha(\tau)^\top z_t^m(\zeta)}(u_t - \alpha_0(\tau)^\top z_t^m(\zeta_0))d\nu_t - \tau\int_{\alpha(\tau)^\top z_t^m(\zeta)}^{\alpha_0(\tau)^\top z_t^m(\zeta_0)}(u_t - \alpha_0(\tau)^\top z_t^m(\zeta_0))d\nu_t - \tau\int_{\alpha_0(\tau)^\top z_t^m(\zeta_0)}^{+\infty}(u_t - \alpha_0(\tau)^\top z_t^m(\zeta_0))d\nu_t \\
&=(1-\tau)\int_{-\infty}^{\alpha(\tau)^\top z_t^m(\zeta)}(\alpha(\tau)^\top z_t^m(\zeta) - \alpha_0(\tau)^\top z_t^m(\zeta_0))d\nu_t + \Omega_1 + \tau\int_{\alpha_0(\tau)^\top z_t^m(\zeta_0)}^{+\infty}(\alpha_0(\tau)^\top z_t^m(\zeta_0) - \alpha(\tau)^\top z_t^m(\zeta))d\nu_t \\
&\geqslant(1-\tau)\int_{-\infty}^{\alpha_0(\tau)^\top z_t^m(\zeta_0)}(\alpha(\tau)^\top z_t^m(\zeta) - \alpha_0(\tau)^\top z_t^m(\zeta_0))d\nu_t + \tau\int_{\alpha_0(\tau)^\top z_t^m(\zeta_0)}^{+\infty}(\alpha_0(\tau)^\top z_t^m(\zeta_0) - \alpha(\tau)^\top z_t^m(\zeta))d\nu_t \qquad (20)\\
&=\alpha_0(\tau)^\top z_t^m(\zeta_0) - \alpha(\tau)^\top z_t^m(\zeta)\left[(1-\tau)\int_{-\infty}^{\alpha_0(\tau)^\top z_t^m(\zeta_0)} d\nu_t + \tau\int_{\alpha_0(\tau)^\top z_t^m(\zeta_0)}^{+\infty} d\nu_t\right] \\
&= \left(\alpha_0(\tau)^\top z_t^m(\zeta_0) - \alpha(\tau)^\top z_t^m(\zeta)\right)\left[\tau - \nu_t(-\infty, F^{-1}_{u_t|\mathcal{F}_{t-1}}(\tau)]\right], \qquad (21)
\end{aligned}
$$

where we use the fact that $\nu_t(-\infty, F^{-1}_{u_t|\mathcal{F}_{t-1}}(\tau)] + \nu_t(F^{-1}_{u_t|\mathcal{F}_{t-1}}(\tau), +\infty) = 1$ (ν_t is a probability measure) and where the inequality in equation (20) follows from the fact that

$$
\Omega_1 := \int_{\alpha(\tau)^\top z_t^m(\zeta)}^{\alpha_0(\tau)^\top z_t^m(\zeta_0)}(u_t - \tau\alpha(\tau)^\top z_t^m(\zeta) - (1-\tau)\alpha_0(\tau)^\top z_t^m(\zeta_0))d\nu_t \geqslant (1-\tau)\int_{\alpha(\tau)^\top z_t^m(\zeta)}^{\alpha_0(\tau)^\top z_t^m(\zeta_0)}(\alpha(\tau)^\top z_t^m(\zeta) - \alpha_0(\tau)^\top z_t^m(\zeta_0))d\nu_t.
$$

Similarly for $\alpha_0(\tau)^\top z_t^m(\zeta_0) < \alpha(\tau)^\top z_t^m(\zeta)$, it holds

$$
\begin{aligned}
&=(1-\tau)\int_{-\infty}^{\alpha_0(\tau)^\top z_t^m(\zeta_0)}(\alpha(\tau)^\top z_t^m(\zeta) - \alpha_0(\tau)^\top z_t^m(\zeta_0))d\nu_t - \Omega_2 + \tau\int_{\alpha(\tau)^\top z_t^m(\zeta)}^{+\infty}(\alpha_0(\tau)^\top z_t^m(\zeta_0) - \alpha(\tau)^\top z_t^m(\zeta))d\nu_t \\
&\geqslant(1-\tau)\int_{-\infty}^{\alpha_0(\tau)^\top z_t^m(\zeta_0)}(z_t^m(\zeta)\alpha(\tau)^\top - z_t^m(\zeta_0)\alpha_0(\tau)^\top)d\nu_t + \tau\int_{\alpha_0(\tau)^\top z_t^m(\zeta_0)}^{+\infty}(\alpha_0(\tau)^\top z_t^m(\zeta_0) - \alpha(\tau)^\top z_t^m(\zeta))d\nu_t \\
&= \left(\alpha(\tau)^\top z_t^m(\zeta) - \alpha_0(\tau)^\top z_t^m(\zeta_0)\right)\left[\tau - \nu_t(-\infty, F^{-1}_{u_t|\mathcal{F}_{t-1}}(\tau)]\right], \qquad (22)
\end{aligned}
$$

where we use

$$
\Omega_2 := \int_{\alpha_0(\tau)^\top z_t^m(\zeta_0)}^{\alpha(\tau)^\top z_t^m(\zeta)}(u_t - \tau\alpha(\tau)^\top z_t^m(\zeta) - (1-\tau)\alpha_0(\tau)^\top z_t^m(\zeta_0))d\nu_t \leqslant \tau\int_{\alpha_0(\tau)^\top z_t^m(\zeta_0)}^{\alpha(\tau)^\top z_t^m(\zeta)}(\alpha_0(\tau)^\top z_t^m(\zeta_0) - \alpha(\tau)^\top z_t^m(\zeta))d\nu_t.
$$

Then by the definition of the τ^{th} quantile, $\nu_t(-\infty, F^{-1}_{u_t|\mathcal{F}_{t-1}}(\tau)] \leqslant \tau$ and the final expressions in both equation (21) and equation (22) are thus non-negative. Thus the expectation in (19) with respect to the measure of z_t is zero for parameters other than the true parameter if and only if it holds for all z_t that $\left(\alpha_0(\tau)^\top z_t^m(\zeta_0) - \alpha(\tau)^\top z_t^m(\zeta)\right) =$

0. The parameters are thus identified if the following identification statement holds:

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\alpha}(\tau)^{\mathsf T} z_t^m(\zeta) - \boldsymbol{\alpha}_0(\tau)^{\mathsf T} z_t^m(\zeta_0)\right] &= \mathbb{E}\left[(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)^{\mathsf T} F_\varepsilon^{-1}(\tau) z_t^m(\zeta) + \boldsymbol{\alpha}_0^{\Delta} F_\varepsilon^{-1}(\tau) z_t^m(G_t(\zeta) - G_t(\zeta_0))\right] \\ &= F_\varepsilon^{-1}(\tau)\mathbb{E}\left[\begin{pmatrix}\boldsymbol{\alpha}-\boldsymbol{\alpha}_0 \\ \boldsymbol{\alpha}^{\Delta}\end{pmatrix}^{\mathsf T}\begin{pmatrix} z_t^m(\zeta) \\ z_t^m(G_t(\zeta)-G_t(\zeta_0))\end{pmatrix}\right] = 0\end{aligned}$$

if and only if $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ and $\zeta = \zeta_0$. This however follows from the global identification Assumption 6, which states that $\mathbb{E}[G_t(\zeta)\, z_t^m, (1-G_t(\zeta))z_t^m, z_t^m(G_t(\zeta)-G_t(\zeta_0))]^{\mathsf T}[G_t(\zeta)\, z_t^m, (1-G_t(\zeta))z_t^m, z_t^m(G_t(\zeta)-G_t(\zeta_0))]$ has full rank for any $\zeta \neq \zeta_0$. Thus $\boldsymbol{\alpha}(\tau) = \boldsymbol{\alpha} F_\varepsilon^{-1}(\tau)$ is identified for an arbitrary $\tau \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1)$. Note that the result does not hold for $\tau = \frac{1}{2}$ for which $F_\varepsilon^{-1}(\tau)$ is zero.

Given this identification result, the consistency for a fixed dimension m would follow once the sample objective function is shown to converge uniformly in probability to its (continuous) population counterpart, which we just analysed above (Theorem 2.1, p. 2121 in Newey & McFadden, 1994); note that for $\tau \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1)$ the parameter space $\boldsymbol{\Theta}_1^\tau$ is a compact subset of $\mathbb{R}^{2(m+1)+2}$. An $m \to \infty$, the consistency of the proposed sieve estimator is obtained by Theorem 1 in Chen & Shen (1998, p. 297), for which we have to check conditions A.1-A.4 therein. Let us first denote $m = \rho_\tau \circ \omega_t$, where $\omega_t : (\boldsymbol{\alpha}^{I}(\tau), \boldsymbol{\alpha}^{II}(\tau), \zeta) \mapsto z_t^{m,\mathsf T}\boldsymbol{\alpha}^{II}(\tau) + (\boldsymbol{\alpha}^{I}(\tau) - \boldsymbol{\alpha}^{II}(\tau))^{\mathsf T} z_t^m G(\xi_t, \zeta, \eta)$, $\boldsymbol{\alpha}(\tau) = [\boldsymbol{\alpha}^{I}(\tau, T), \boldsymbol{\alpha}^{II,T}(\tau)]^{\mathsf T}$ and $\zeta = [\zeta, \eta]^{\mathsf T}$. For condition A.4, we have to show that the function $m = \rho_\tau \circ \omega_t$ is Lipschitz. By Assumption 1, the transition function $G_t(\zeta)$ is Lipschitz in ζ, and by construction, ω_t is Lipschitz in $\boldsymbol{\alpha}$ and $G()$. Thus, ω_t is Lipschitz in $\boldsymbol{\alpha}$ and ζ since the property is preserved under function composition.[19] The piecewise linear function ρ_τ is Lipschitz as well and so is thus $m = \rho_\tau \circ \omega_t$. Further, note that we actually have a special case of $s = 1$ in the Hölder condition A.4, which by Chen & Shen (1998, Remark 1(c)) implies their condition A.2. In addition to this, condition A.1 holds by Assumption 5 which states that u_t is β-mixing with decay rate satisfying $\beta_s \leqslant \beta_0 s^{-(2+\delta)}$ for some $\delta > 0$. Finally, we note that, by Chen (2008, p. 5595), it holds for Lipschitz functions that the bracketing numbers $\log N_{[]}(\epsilon^s, \mathcal{F}_n, \|\cdot\|_2) \leqslant \log N(\epsilon, \Theta_1^\tau, \|\cdot\|) \leqslant Cm\log(\frac{1}{\epsilon})$ for some constant $C > 0$, where m is the dimension of the sieve parameter space, and that this implies their condition A.3. Therefore, we can apply Theorem 1 in Chen & Shen (1998) and conclude that the first-stage sieve estimator is consistent.

Next to derive the convergence rate of the first-stage estimator, we have to show that the directional derivative of the non-differentiable g_n around the true value $\mathbf{a}_0$ is positive in every direction with probability tending to one. Let $\varphi_t(\mathbf{a}) = \sum_{k=1}^{K} \mathbf{x}_t(\mathbf{a})(\tau_k - \mathbb{1}\left\{u_t \leqslant \mathbf{a}^{\mathsf T}\mathbf{x}_t(\mathbf{a})\right\})$. Then we have to prove that

$$\forall \epsilon > 0 : \exists B < \infty : \lim_{n\to\infty} \mathbf{P}\left[\inf_{\boldsymbol{\lambda}\in\mathbb{R}^{2(m+1)+K+2}:\|\boldsymbol{\lambda}\|=1} \mathbb{P}_n \boldsymbol{\lambda}^{\mathsf T}\varphi_t\left(\mathbf{a}_0 + B\left(\frac{m}{n}\right)^{\frac{1}{2}}\boldsymbol{\lambda}\right) > 0\right] > 1-\epsilon. \quad (23)$$

Adding and subtracting $\mathbb{P}_n\boldsymbol{\lambda}^{\mathsf T}\mathbb{E}\left[\varphi_t\left(\mathbf{a}_0 + B\left(\frac{m}{n}\right)^{\frac{1}{2}}\boldsymbol{\lambda}\right)\middle|\mathcal{F}_{t-1}\right] - \mathbb{P}_n\boldsymbol{\lambda}^{\mathsf T}\mathbb{E}\left[\varphi_t(\mathbf{a}_0)|\mathcal{F}_{t-1}\right]$ as well as $\mathbb{P}_n\boldsymbol{\lambda}^{\mathsf T}\varphi_t(\mathbf{a}_0)$,

[19] If both f and g are Lipschitz so is $f \circ g$ since $(f\circ g)(x) - (f\circ g)(x_0) = f(g(x)) - f(g(x_0)) \leqslant C_f(g(x)-g(x_0)) \leqslant C_f C_g(x - x_0)$ for finite constants C_f and C_g.

the random quantity in (23) can be rewritten as:

$$
\begin{aligned}
&\mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \varphi_t \left(\boldsymbol{a}_0 + B \left(\frac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda} \right) && (24)\\
&\quad = \left\{ \mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \varphi_t \left(\boldsymbol{a}_0\right) \right\} && (25)\\
&\quad + \left\{ \left(\mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \varphi_t \left(\boldsymbol{a}_0 + B \left(\frac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda} \right) - \mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \varphi_t \left(\boldsymbol{a}_0\right) \right) \right. && (26)\\
&\quad \left. - \left(\mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \mathbb{E} \left[\varphi_t \left(\boldsymbol{a}_0 + B \left(\frac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda} \right) \middle| \mathcal{F}_{t-1} \right] - \mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \mathbb{E} \left[\varphi_t \left(\boldsymbol{a}_0\right) \middle| \mathcal{F}_{t-1} \right] \right) \right\} \\
&\quad + \left\{ \mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \mathbb{E} \left[\varphi_t \left(\boldsymbol{a}_0 + B \left(\frac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda} \right) \middle| \mathcal{F}_{t-1} \right] - \mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \mathbb{E} \left[\varphi_t \left(\boldsymbol{a}_0\right) \middle| \mathcal{F}_{t-1} \right] \right\}. && (27)
\end{aligned}
$$

This expansion can be now analysed term by term. Term (26) will turn out to be stochastically negligible, whereas term (25) and (27) can be made explicit. Let us start by rewriting (27) in the following way, using the definition of φ_t:

$$
\begin{aligned}
&\mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \mathbb{E} \left[\varphi_t \left(\boldsymbol{a}_0 + B \left(\frac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda} \right) \middle| \mathcal{F}_{t-1} \right] - \mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \mathbb{E} \left[\varphi_t \left(\boldsymbol{a}_0\right) \middle| \mathcal{F}_{t-1} \right] \\
&\quad = \mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \mathbb{E} \left[\sum_{k=1}^{K} \left(\tau_k - \mathbb{1} \left\{ u_t \leqslant \left(\boldsymbol{a}_0 + B \left(\frac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda} \right)^\mathsf{T} \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \right\} \right) \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \middle| \mathcal{F}_{t-1} \right] \\
&\quad - \mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \mathbb{E} \left[\sum_{k=1}^{K} \left(\tau_k - \mathbb{1} \left\{ u_t \leqslant \boldsymbol{a}_0^\mathsf{T} \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \right\} \right) \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \middle| \mathcal{F}_{t-1} \right] \\
&\quad = -\mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \sum_{k=1}^{K} \left(F_{u_t|\mathcal{F}_{t-1}} \left(\left(\boldsymbol{a}_0 + B \left(\frac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda} \right)^\mathsf{T} \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \right) - F_{u_t|\mathcal{F}_{t-1}} \left(\boldsymbol{a}_0^\mathsf{T} \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \right) \right) \boldsymbol{x}_{t,k}(\boldsymbol{a}_0).
\end{aligned}
$$

Applying the Taylor expansion around $\boldsymbol{a}_0$ to the first term for each $t \in \mathcal{I}_{m,n}$ yields:

$$
\begin{aligned}
&- \mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \sum_{k=1}^{K} f_{u_t|\mathcal{F}_{t-1}} \left(\boldsymbol{a}_0^\mathsf{T} \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \right) \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \boldsymbol{x}_{t,k}(\boldsymbol{a}_0)^\mathsf{T} B \left(\frac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda} + \mathcal{O}_p \left(m^1 n^{-1} \right) \\
&= -B \left(\frac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda}^\mathsf{T} \mathbb{P}_n \left[\sum_{k=1}^{K} \frac{1}{\sigma_t} f_\varepsilon \left(F_\varepsilon^{-1}(\tau_k) \right) \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \boldsymbol{x}_{t,k}(\boldsymbol{a}_0)^\mathsf{T} \right] \boldsymbol{\lambda} + \mathcal{O}_p \left(m^1 n^{-1} \right) \\
&= \left(\frac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda}^\mathsf{T} \mathbf{D}_{1,n,m} \boldsymbol{\lambda} + \mathcal{O}_p \left(m^1 n^{-1} \right),
\end{aligned}
$$

using equations (6), (4) for the last step and defining $\mathbf{D}_{1,n,m} =$

$$
-\mathbb{E}\mathbb{P}_n \frac{1}{\sigma_t} \left[\begin{array}{c:c:c}
\boldsymbol{\iota}_K^\mathsf{T} h(\mathbf{q}) \nu(\zeta) \nu(\zeta)^\mathsf{T} \otimes \boldsymbol{z}_t \boldsymbol{z}_t^\mathsf{T} & \boldsymbol{z}_t^\mathsf{T} \bar{\alpha}(\zeta) (\nu(\zeta) \otimes \boldsymbol{z}_t) \otimes h(\boldsymbol{\iota}_K)^\mathsf{T} & \boldsymbol{\iota}_K^\mathsf{T} h(\mathbf{q}) \boldsymbol{\alpha}^\Delta \boldsymbol{z}_t (\nu(\zeta) \otimes \boldsymbol{z}_t) \frac{\partial G}{\partial \zeta^\mathsf{T}} \\ \hdashline
 & \mathrm{diag}(\mathbf{s}) (\bar{\boldsymbol{\alpha}}^\mathsf{T} \boldsymbol{z}_t)^2 & \bar{\boldsymbol{\alpha}}^\mathsf{T} \boldsymbol{z}_t(\zeta) \boldsymbol{\alpha}^\Delta \boldsymbol{z}_t h(\boldsymbol{\iota}_K) \frac{\partial G}{\partial \zeta^\mathsf{T}} \\ \hdashline
 & & \boldsymbol{\iota}_K^\mathsf{T} h(\mathbf{q}) (\boldsymbol{\alpha}^\Delta \boldsymbol{z}_t)^2 \frac{\partial G}{\partial \zeta} \frac{\partial G}{\partial \zeta^\mathsf{T}}
\end{array} \right]
$$

where $\bar{\boldsymbol{\alpha}} = G_t(\zeta)\, \boldsymbol{\alpha}^{\mathrm{I}} + (1 - G_t(\zeta)) \boldsymbol{\alpha}^{\mathrm{II}}$, $\nu(\zeta) = [G_t(\zeta), 1 - G_t(\zeta)]^\mathsf{T}$, $\mathbf{s} = (s_1, \ldots, s_K)^\mathsf{T}$, $\mathbf{q} = (q_1, \ldots, q_K)^\mathsf{T}$, $h(\boldsymbol{\chi}) = \mathbf{s} \odot \mathbf{q} \odot \boldsymbol{\chi}$ and where $\odot$ is the Hadamard product.

To analyse the remaining terms, let $\eta_t(\nu) = \varphi_t(\boldsymbol{a}_0 + \boldsymbol{\nu}) - \varphi_t(\boldsymbol{a}_0)$. Then term (26) is negligible in probability with rate $\left(\frac{m}{n}\right)^{-\frac{1}{2}}$ if

$$
\sup_{\|\nu\| \leqslant B \left(\frac{m}{n}\right)^{\frac{1}{2}}} \left| \mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \left(\eta_t(\nu) - \mathbb{E} \left[\eta_t(\nu) \middle| \mathcal{F}_{t-1} \right] \right) \right| = o_p \left(\frac{1}{\sqrt{mn}} \right).
$$

The considered process is a martingale difference sequence. The next step is to divide the ball defined as $\left\{\boldsymbol{\nu} \in \mathbb{R}^{2(m+1)+K+2} : \|\boldsymbol{\nu}\| \leqslant B(\frac{m}{n})^{\frac{1}{2}}\right\}$ in equation (23) into cubes $\mathcal{C}_j \subset \mathbb{R}^{2(m+1)+K+2}$ centred at ν_j and with side-length $m^{\frac{1}{2}} n^{-\frac{5}{2}}$. The resulting cardinality for $2(m+1)+K+2$ dimensions is then $N(n) := \|\{\mathcal{C}_j\}\| = (2n)^{2(m+1)+K+2}$. Now then for each $k \in \mathcal{I}_{1,K}$, the term $\eta_t(\nu)$ can be bounded by

$$\eta_t(\nu) \leqslant \eta_t(\nu_j) + b_{k,t}(\nu_j)\mathbf{x}_{t,k}(\mathbf{a}_0), \tag{28}$$

and similarly,

$$\eta_t(\nu) \geqslant \eta_t(\nu_j) + (b_{k,t}(\nu_j) - d_{k,t}(\nu_j))\,\mathbf{x}_{t,k}(\mathbf{a}_0), \tag{29}$$

with $b_{k,t}(\nu_j)\mathbf{x}_{t,k}(\mathbf{a}_0)$ and $d_{k,t}(\nu_j)\mathbf{x}_{t,k}(\mathbf{a}_0)$ being the process η_t evaluated at the maximum possible distance on each axis from the centre to the boundary of the cube and the maximum possible distance on each axis between the boundaries of the cube $\mathcal{C}_j$, respectively:

$$\begin{aligned} b_{k,t}(\nu_j) = {} & \mathbb{1}\left\{u_t < (\mathbf{a}_0 + \nu_j)^\mathsf{T}\,\mathbf{x}_{t,k}(\mathbf{a}_0)\right\} \\ & - \mathbb{1}\left\{u_t < (\mathbf{a}_0 + \nu_j)^\mathsf{T}\,\mathbf{x}_{t,k}(\mathbf{a}_0) + B\left(n^{\frac{1}{2}} m^{-\frac{5}{2}}\right)\|\mathbf{x}_{t,k}(\mathbf{a}_0)\|\right\}, \end{aligned}$$

$$\begin{aligned} d_{\tau,t}(\nu_j) = {} & \mathbb{1}\left\{u_t < (\mathbf{a}_0 + \nu_j)^\mathsf{T}\,\mathbf{x}_{t,k}(\mathbf{a}_0) + B\left(n^{\frac{1}{2}} m^{-\frac{5}{2}}\right)\|\mathbf{x}_{t,k}(\mathbf{a}_0)\|\right\} \\ & - \mathbb{1}\left\{u_t < (\mathbf{a}_0 + \nu_j)^\mathsf{T}\,\mathbf{x}_{t,k}(\mathbf{a}_0) - B\left(n^{\frac{1}{2}} m^{-\frac{5}{2}}\right)\|\mathbf{x}_{t,k}(\mathbf{a}_0)\|\right\}. \end{aligned}$$

Taking expectations of (29) and subtracting it from (28) implies that, for all $\nu \in \mathcal{C}_j$, for all t, and for all k, it holds that

$$\begin{aligned} (\eta_t(\nu) - \mathbb{E}[\eta_t(\nu)|\mathcal{F}_{t-1}]) \leqslant {} & (\eta(\nu_j) - \mathbb{E}[\eta_t(\nu_j)|\mathcal{F}_{t-1}]) \\ & + (b_{k,t}(\nu_j) - \mathbb{E}[b_{k,t}(\nu_j)|\mathcal{F}_{t-1}])\,\mathbf{x}_{t,k}(\mathbf{a}_0) + \mathbb{E}[d_{k,t}(\nu_t)|\mathcal{F}_{t-1}]\,\mathbf{x}_{t,k}(\mathbf{a}_0), \end{aligned}$$

which implies that

$$\sup_{\|\nu\| \leqslant B\left(\frac{m}{n}\right)^{\frac{1}{2}}} \left|\mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} (\eta_t(\nu) - \mathbb{E}[\eta_t(\nu)|\mathcal{F}_{t-1}])\right|$$

$$\leqslant \max_{j \in \mathcal{I}_{1,N(n)}} \left|\mathbb{P}_n \left\|\boldsymbol{\lambda}^\mathsf{T}\mathbf{x}_{t,k}(\mathbf{a}_0)\right\| (b_{k,t}(\nu_j) - \mathbb{E}[b_{k,t}(\nu_j)|\mathcal{F}_{t-1}])\right| \tag{30}$$

$$+ \max_{j \in \mathcal{I}_{1,N(n)}} \left|\mathbb{P}_n \left\|\boldsymbol{\lambda}^\mathsf{T}\mathbf{x}_{t,k}(\mathbf{a}_0)\right\| \mathbb{E}[d_{k,t}(\nu_t)|\mathcal{F}_{t-1}]\right| \tag{31}$$

$$+ \max_{j \in \mathcal{I}_{1,N(n)}} \left|\mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} (\eta(\nu_j) - \mathbb{E}[\eta_t(\nu_j)|\mathcal{F}_{t-1}])\right|. \tag{32}$$

Expressions (30), (31), and (32) are equivalent to expressions (A.5)–(A.7) in Xiao & Koenker (2009), who show these terms are asymptotically negligible of order $\sqrt{m/n}$ in probability. As their proof is general and relies on the existence of the exponential bound on innovations imposed in Assumption 8, the results also apply to the present analysis since the current problem is piecewise linear with a bounded transition function, with finite second moments of the conditional volatility process, and satisfying Assumption 8.

Hence, with term (26) being negligible, equation (24) can be written as

$$\mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \varphi_t\left(\mathbf{a}_0 + B\left(\frac{m}{n}\right)^{\frac{1}{2}}\boldsymbol{\lambda}\right) = \mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \varphi_t(\mathbf{a}_0) + B\left(\frac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda}^\mathsf{T}\mathbf{D}_{1,n,m}\boldsymbol{\lambda} + o_p\left(\left(\frac{m}{n}\right)^{\frac{1}{2}}\right). \tag{33}$$

Whenever the right hand side of this equation exceeds zero, it is implied that so does the left hand side. The left-hand side is however positive with probability tending to one as $B \to \infty$ and $n \to \infty$ since the following equation holds by Assumption 6:

$$\inf_{\lambda \in \mathbb{R}^{2(m+1)+K+2}:\|\lambda\|=1} \left(\frac{m}{n}\right)^{-\frac{1}{2}} \mathbb{P}_n \boldsymbol{\lambda}^T \varphi_t(\mathbf{a}_0) > -\frac{B}{2}\lambda_{n,\min} - o_p(1) < 0,$$

noting that $\mathbb{P}_n \boldsymbol{\lambda}^T \varphi_t(\mathbf{a}_0) = \mathcal{O}_p(\sqrt{m/n})$ and $\lambda_{n,\min} > 0$ as $n \to \infty$. Statement (23) and the convergence rate following from it are thus verified. □

Proof of Theorem 2: Let $\hat{\nu} = \hat{\mathbf{a}}_n - \mathbf{a}_0$ where $\hat{\mathbf{a}}_n$ solves the objective function defined in equation (9). By Theorem 1, we can write $\hat{\nu}$ as $B\left(\frac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda}$, B in a compact set, with a probability arbitrarily close to 1. This substitution in equation (33) leads to

$$\mathbb{P}_n \boldsymbol{\lambda}^T \varphi_t(\hat{\mathbf{a}}_n) = \mathbb{P}_n \boldsymbol{\lambda}^T \varphi_t(\mathbf{a}_0) + \boldsymbol{\lambda}^T \mathbf{D}_{1,n,m}(\hat{\mathbf{a}}_n - \mathbf{a}_0) + o_p\left(\left(\frac{m}{n}\right)^{\frac{1}{2}}\right).$$

By construction, the moment function on the left-hand side is zero at the estimate $\hat{\mathbf{a}}_n$. Thus, the right hand side satisfies for all $\lambda \in \mathbb{R}^m$, $\|\lambda\| = 1$,

$$\boldsymbol{\lambda}^T \left[\mathbb{P}_n \varphi_t(\mathbf{a}_0) + \mathbf{D}_{1,n,m}(\hat{\mathbf{a}}_n - \mathbf{a}_0) + o_p\left(\frac{1}{\sqrt{n}}\right)\right] = 0,$$

and hence, the expression inside the bracket must be zero. After pre-multiplying it by $\sqrt{n}$ and $\mathbf{D}_{1,n,m}$, the Bahadur representation for $\sqrt{n}(\hat{\mathbf{a}}_n - \mathbf{a}_0)$ follows as well as the one for $\sqrt{n}(\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0)$ by only considering the first $2(m+1)$ and the last 2 elements. The submatrix consisting of the corner blocks (upper right, upper left, bottom right, and bottom left corners) of the matrix $\mathbf{D}_{1,n,m}$ is denoted $\mathbf{D}_m$ here.

Additionally, applying the central limit theorem (Theorem 18.5.3 in Ibragimov & Linnik, 1971) to $\frac{1}{\sqrt{n}}\sum_{t=1}^N \varphi_t(\mathbf{a}_0)$, it follows for any linear combination of the components of the $(2(m+1)+2)$-dimensional Bahadur representation that

$$\sqrt{n}\boldsymbol{\mu}_n^T(\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0) \rightsquigarrow \mathcal{N}\left(0\,,\ \frac{\sum_{k=1}^{K}\sum_{k'=1}^{K} q_k q_{k'}(\tau_k \wedge \tau_{k'})(1-\tau_k \vee \tau_{k'})}{\left(\sum_{k=1}^{K} q_k^2 f_\varepsilon\left(F_\varepsilon^{-1}(\tau_k)\right)\right)^2} \lim_{m\to\infty} \boldsymbol{\mu}_m^T \mathbf{D}_m^{-1} \boldsymbol{\mu}_m\right) \tag{34}$$

where $\boldsymbol{\mu}_m \in \mathbb{R}^{2(m+1)+2}$ is such that the limit $\lim_{m\to\infty} \boldsymbol{\mu}_m^T \mathbf{D}_m^{-1} \boldsymbol{\mu}_m$ exists. The assumptions of the central limit theorem are satisfied due to the moment and mixing conditions stated in Assumption 5, which ensure that the $(2+\delta)$ moments of the data exist and mixing coefficient satisfying $\beta_s \to 0$ and $\sum_{s=1}^{\infty} \beta_s^{\delta/(2+\delta)} < +\infty$. □

For the second stage, the directional derivative of the objective function as in equation (11) is re-defined as

$$g_n(\theta, \boldsymbol{\alpha}) = g_n\left((\theta^I, \theta^{II}, \zeta), (\boldsymbol{\alpha}^I, \boldsymbol{\alpha}^{II}, \zeta)\right) := \mathbb{P}_n \begin{bmatrix} z_t(\boldsymbol{\alpha})G_t(\zeta) \\ z_t(\boldsymbol{\alpha})(1-G_t(\zeta)) \\ \theta^{\Delta T} z_t(\boldsymbol{\alpha}) \frac{\partial G_t(\zeta)}{\partial \zeta} \end{bmatrix} \left(\tau - \mathbb{1}\left\{u_t \leqslant [\theta^I, \theta^{II}]^T z_t(\boldsymbol{\alpha}, \zeta)\right\}\right)$$

and the one for the population, using the law of iterative expectations, as

$$g(\theta,\boldsymbol{\alpha}) = g\left((\theta^{I},\theta^{II},\zeta),(\boldsymbol{\alpha}^{I},\boldsymbol{\alpha}^{II},\zeta)\right) := \mathbb{E}\left[\begin{bmatrix} z_t(\boldsymbol{\alpha})G_t(\zeta) \\ z_t(\boldsymbol{\alpha})(1-G_t(\zeta)) \\ \theta^{\Delta T} z_t(\boldsymbol{\alpha})\frac{\partial G_t(\zeta)}{\partial \zeta} \end{bmatrix}\left(\tau - F_{u|\mathcal{F}_{t-1}}\left([\theta^{I},\theta^{II}]^{T} z_t(\boldsymbol{\alpha},\zeta)\right)\right)\right].$$

Note that we re-estimate ζ, and therefore, we consider only the first-order conditions for the parameter ζ that is a part of θ; parameters within $\boldsymbol{\alpha}$ are fixed and will be substituted for by the first-stage estimates. Due to Assumption 3, the population derivative $g(\theta,\boldsymbol{\alpha})$ is differentiable and the partial derivatives with respect to the two parameter vectors θ and $\boldsymbol{\alpha}$ evaluated at the true parameters are given by $\Gamma_{\theta,0} = \Gamma_\theta(\theta_0,\boldsymbol{\alpha}_0;\mathbf{P})$, and $\Gamma_{\alpha,m,0} := \Gamma_{\alpha,m}(\theta_0,\boldsymbol{\alpha}_0;\mathbf{P})$ with

$$\Gamma_\theta(\theta,\boldsymbol{\alpha};\mu) := -\frac{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)}{\sigma_\varepsilon}\times$$

$$\int \left[\begin{array}{c:c} \begin{bmatrix} G_t(\zeta)^2 & G_t(\zeta)(1-G_t(\zeta)) \\ G_t(\zeta)(1-G_t(\zeta)) & (1-G_t(\zeta))^2 \end{bmatrix} \otimes z_t(\boldsymbol{\alpha})z_t(\boldsymbol{\alpha})^{T} & \theta^{\Delta T} z_t(\boldsymbol{\alpha})\begin{bmatrix} G_t(\zeta) \\ 1-G_t(\zeta) \end{bmatrix}\otimes z_t(\boldsymbol{\alpha})\frac{\partial G_t(\zeta)}{\partial \zeta^{T}} \\ \hdashline & \left(\theta^{\Delta T} z_t(\boldsymbol{\alpha})\right)^2 \frac{\partial G_t(\zeta)}{\partial \zeta}\frac{\partial G_t(\zeta)}{\partial \zeta^{T}} \end{array}\right] d\mu$$

and

$$\Gamma_{\alpha,m}(\theta,\boldsymbol{\alpha};\mu) := -\frac{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)}{\sigma_\varepsilon}\int \begin{bmatrix} z_t(\boldsymbol{\alpha})G_t(\zeta) \\ z_t(\boldsymbol{\alpha})(1-G_t(\zeta)) \\ \theta^{\Delta T} z_t(\boldsymbol{\alpha})\frac{\partial G_t(\zeta)}{\partial \zeta} \end{bmatrix}\theta^{T}\left[\begin{matrix}\begin{bmatrix} G_t(\zeta) \\ 1-G_t(\zeta) \end{bmatrix} \otimes \begin{bmatrix} 0_{p+1,m} \\ [L^1,\ldots,L^q]^{T}\otimes z_t^{m} \end{bmatrix} \\ \iota_q^{T}[L^1,\ldots,L^q]^{T}\otimes z_t^{mT}\boldsymbol{\alpha}^{\Delta}\frac{\partial G}{\partial \zeta}\end{matrix}\right] d\mu,$$

respectively, where L is the lag operator. These expectations are well-defined and exist by Assumptions 1–5. In addition to this, $\Gamma_{\theta,0}$ is positive definite and has full rank due to the invertibility of the conditional scale process in Assumption 2 and Assumption 9. Finally, let $\Gamma_{\theta,n}(\theta,\boldsymbol{\alpha}) = \Gamma_\theta(\theta,\boldsymbol{\alpha};\mathbb{P}_n)$ and $\Gamma_{\alpha,n}(\theta,\boldsymbol{\alpha}) := \Gamma_{\alpha,m}(\theta,\boldsymbol{\alpha};\mathbb{P}_n)$ be the corresponding sample analogues.

Proof of Theorem 3: It needs to be shown that $\left\|\widehat{\theta}_n(\tau)-\theta_0(\tau)\right\| = \mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$. For this, note that $g(\theta,\boldsymbol{\alpha})$ is differentiable for any θ. Thus, the first-order Taylor expansion around $\theta_0(\tau)$ can be applied, and due to continuity of Γ_θ in θ, it follows that

$$g\left(\widehat{\theta}_n(\tau),\boldsymbol{\alpha}_0\right) - g(\theta_0(\tau),\boldsymbol{\alpha}_0) = (\Gamma_{\theta,0}+o_p(1))\left(\widehat{\theta}_n(\tau)-\theta_0(\tau)\right) + .$$

Taking norms, a bound for the right hand side is obtained with a probability arbitrarily close to 1 for $n\to\infty$:

$$\left\|g\left(\widehat{\theta}_n(\tau),\boldsymbol{\alpha}_0\right) - g(\theta_0(\tau),\boldsymbol{\alpha}_0)\right\| \geqslant \frac{1}{2}\lambda_{\min}(\Gamma_{\theta,0})\left\|\left(\widehat{\theta}_n(\tau)-\theta_0(\tau)\right)\right\|, \tag{35}$$

with $\lambda_{\min}(\Gamma_{\theta,0})$ being the the smallest eigenvalue of $\Gamma_{\theta,0}$, which is strictly positive as argued above. Since $g(\theta_0(\tau),\boldsymbol{\alpha}_0) = 0$, it is sufficient to show that $\left\|g\left(\widehat{\theta}_n(\tau),\boldsymbol{\alpha}_0\right)\right\| = \mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$ to prove the theorem. Using the triangle inequality, it follows that

$$
\begin{aligned}
\left\|g\left(\widehat{\theta}_n(\tau),\boldsymbol{\alpha}_0\right)\right\| &\leqslant \left\|g\left(\widehat{\theta}_n(\tau),\boldsymbol{\alpha}_0\right)-g\left(\widehat{\theta}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)\right\|+\left\|g\left(\widehat{\theta}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)\right\| \\
&\leqslant \left\|g\left(\widehat{\theta}_n(\tau),\boldsymbol{\alpha}_0\right)-g\left(\widehat{\theta}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)\right\| && (36)\\
&+\left\|g\left(\widehat{\theta}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)-g\left(\theta_0(\tau),\boldsymbol{\alpha}_0\right)-g_n\left(\widehat{\theta}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)+g_n\left(\theta_0(\tau),\boldsymbol{\alpha}_0\right)\right\| && (37)\\
&+\left\|g_n\left(\widehat{\theta}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)\right\| && (38)\\
&+\left\|g_n\left(\theta_0(\tau),\boldsymbol{\alpha}_0\right)\right\|, && (39)
\end{aligned}
$$

where $g\left(\theta_0(\tau),\boldsymbol{\alpha}_0\right)=0$ was subtracted within the second norm (37). By the central limit theorem (Theorem 18.5.3 in Ibragimov & Linnik, 1971), the existence of the $(2+\delta)$ moments of z_t^m, $z_t^m G_t(\zeta_0)$, and $z_t^m \partial G_t(\zeta)/\partial\zeta$, respectively, and the boundedness of both $G_t(\zeta_0)$ and $\partial G_t(\zeta_0)/\partial\zeta$ implies that the expression (39) is tight and it holds $\left\|g_n\left(\theta_0(\tau),\boldsymbol{\alpha}_0\right)\right\|=\mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$. The remaining terms (36), (37), and (38) can again be analysed separately. Starting with the first term, again using the triangle inequality, and changing the signs within the norm, term (36) can be bounded by

$$
\begin{aligned}
\left\|g\left(\widehat{\theta}_n(\tau),\boldsymbol{\alpha}_0\right)-g\left(\widehat{\theta}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)\right\| &\leqslant \left\|g\left(\widehat{\theta}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)-g\left(\widehat{\theta}_n(\tau),\boldsymbol{\alpha}_0\right)-\Gamma_{\alpha,n}\left(\widehat{\theta}_n(\tau),\boldsymbol{\alpha}_0\right)\left(\widehat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0\right)\right\| && (40)\\
&+\left\|\Gamma_{\alpha,n}\left(\widehat{\theta}_n(\tau),\boldsymbol{\alpha}_0\right)\left(\widehat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0\right)-\Gamma_{\alpha,n}\left(\theta_0(\tau),\boldsymbol{\alpha}_0\right)\left(\widehat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0\right)\right\| && (41)\\
&+\left\|\Gamma_{\alpha,n}\left(\theta_0(\tau),\boldsymbol{\alpha}_0\right)\left(\widehat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0\right)\right\|
\end{aligned}
$$

Applying the Taylor series expansion of $g\left(\widehat{\theta}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)$ around $\boldsymbol{\alpha}_0$ in (40) and using the fact that $\Gamma_{\alpha,n}$ is Lipschitz in $\boldsymbol{\alpha}$ (since $\partial G_t(\zeta_0)/\partial\zeta$ is Lipschitz and σ_t is bounded as it is invertible to an ARCH model by Assumption 2), term (40) is negligible in probability wrt. $\left\|\widehat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0\right\|^2=\mathcal{O}_p\left(m/n\right)=o_p\left(n^{-1/2}\right)$ by Theorem 1 and Assumption 7. Similarly for equation (41), we use that $\Gamma_{\alpha,n}\left(\widehat{\theta}_n(\tau),\boldsymbol{\alpha}_0\right)$ is Lipschitz in $\theta_0(\tau)$, which has bounded parameter space. Thus, (40) reduces to

$$
\begin{aligned}
&\left\|g\left(\widehat{\theta}_n(\tau),\boldsymbol{\alpha}_0\right)-g\left(\widehat{\theta}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)\right\| && (42)\\
&\leqslant \mathcal{O}_p\left(\left\|\widehat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0\right\|^2\right)+\mathcal{O}_p\left(\left\|\widehat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0\right\|\left\|\widehat{\theta}_n(\tau)-\theta_0(\tau)\right\|\right)+\left\|\Gamma_{\alpha,m,0}\left(\widehat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0\right)\right\|\\
&=\left\|\Gamma_{\alpha,m,0}\left(\widehat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0\right)\right\|\left(1+o_p(1)\right)=\mathcal{O}_p\left(n^{-1/2}\right),
\end{aligned}
$$

where the the last term follows from (i) the fact that the elementwise $\boldsymbol{\Gamma}_{\alpha,n}\to\boldsymbol{\Gamma}_{\alpha,0}$ in probability by law of large numbers (the respective moments exist by Assumption 5) and Slutsky's lemma and (ii) equation (34), which applies due to Assumption 9.

In a next step, we analyse the remaining terms (37) and (38), for which we have to check the conditions of Lemma 4.2 in Chen (2008). For this, let

$$
m_\tau(z_t,\theta,\boldsymbol{\alpha})=\begin{bmatrix} z_t(\boldsymbol{\alpha})G_t(\zeta)\\ z_t(\boldsymbol{\alpha})(1-G_t(\zeta))\\ \theta^{\Delta T}z_t(\boldsymbol{\alpha})\frac{\partial G_t(\zeta)}{\partial\zeta}\end{bmatrix}\left(\tau-\mathbb{1}\left\{u_t\leqslant[\theta^{I},\theta^{II}]^{T}z_t(\boldsymbol{\alpha},\zeta)\right\}\right) \tag{43}
$$

so that $g_n(\theta,\boldsymbol{\alpha})=\mathbb{P}_n m_\tau(z_t(\boldsymbol{\alpha}),\theta,\boldsymbol{\alpha})$ and $g(\theta,\boldsymbol{\alpha})=\mathbb{E}m_\tau(z_t,\theta,\boldsymbol{\alpha})$. Then if z_t is stationary, which is true by Assumption 2 and 3, has β-mixing decay rate as in Assumption 5 (see e.g. Carrasco & Chen (2002), Meitz & Saikkonen (2008)), Θ_2^τ is a compact subset of $\mathbb{R}^{2(p+q+1)+2}$ and Θ_1 one of $\mathbb{R}^{2(m+1)}\times\mathbb{R}\times\mathbb{R}_+$, we have to verify for each jth component $\mathfrak{m}_{\tau,j}$, $j \in \mathcal{I}_{2(p+q+1)+2}$, of $\mathfrak{m}_\tau$ that

$$\left(\mathbb{E}\left[\sup_{(\theta'',\zeta'',\alpha',\zeta')\in\mathfrak{U}_\delta((\theta''_0,\zeta_0,\alpha'_0,\zeta_0)}\left|\mathfrak{m}_{\tau,j}(z_t,\theta'',\zeta'',\alpha',\zeta')-\mathfrak{m}_{\tau,j}(z_t,\theta''_0,\zeta_0,\alpha'_0,\zeta_0)\right|^r\right]\right)^{\frac{1}{r}} \leqslant K_j\delta^{s_j},$$

where $\alpha' = [\alpha^{I}, \alpha^{II}]^{\top}$ and $\theta'' = [\theta^{I}, \theta^{II}]^{\top}$, for some s_j that is bounded by the degree of smoothness of $G_t(\zeta)$, for some constant $K_j > 0$, and for $r = 2 + \delta$ satisfying the restriction in Assumption 5 to claim that:

$$\sup_{(\theta,\alpha)\in\mathfrak{U}_\delta(\theta_0(\tau),\alpha_0)}\|g(\theta,\alpha)-g(\theta_0(\tau),\alpha_0)-g_n(\theta,\alpha)+g_n(\theta_0(\tau),\alpha_0)\| = o_p\left(n^{-\frac{1}{2}}\right)$$

with $\mathfrak{U}_\delta := \{(\theta,\alpha) \in \Theta_2^{\tau} \times \Theta_1 : \|\theta-\theta_0(\tau)\| < \delta, \|\alpha-\alpha_0\| < \delta\}$ by Lemma 4.2 in Chen (2008). As discussed in Chen (2008), $\mathfrak{m}_{\tau,j}$ needs to be a member of a function class with covering numbers satisfying condition $\int_0^\infty \sqrt{\log N(\epsilon^{1/s_j}, \mathcal{H}, \|\cdot\|_{\mathcal{H}})}d\epsilon < \infty$, where the degree of smoothness satisfies $d = 1 \geqslant 2/(2s_j)$ with $s_j = 1$ in our case. Alternatively, we can make use of the class of monotone functions which is sufficient for the former condition; for details, see Chen (2008). Consequently, we either need continuity of $\frac{\partial G_t(\zeta)}{\partial\zeta}$ (Theorem 2.7.1 in van der Vaart & Wellner, 1996) or monotonicity (Theorem 2.7.5 in van der Vaart & Wellner, 1996) of $\frac{\partial G_t(\zeta)}{\partial\zeta}$ with respect to ξ_t and Assumption 1 ensures that the transition function belongs to one of these classes.

The uniform boundedness relative to the L^r-norm of the distance between (43) evaluated at any two parameter values within a neighbourhood of the true parameters can be shown as follows.

By definition,

$$z_{t,j}(\theta'',\zeta'',\alpha',\zeta') = \begin{cases} G(\zeta'')\left[z_t(\alpha',\zeta')\right]_j & \text{if} \quad 0 < j \leqslant p+q+1, \\ (1-G(\zeta''))\left[z_t(\alpha',\zeta')\right]_{j-(p+q+1)} & \text{if} \quad p+q+1 < j \leqslant 2(p+q+1), \\ \theta''^{\Delta\top} z_t(\alpha',\zeta')\frac{\partial G(\zeta'',\eta'')}{\partial\zeta} & \text{if} \quad j = 2(p+q+1)+1, \\ \theta''^{\Delta\top} z_t(\alpha',\zeta')\frac{\partial G(\zeta'',\eta'')}{\partial\eta} & \text{if} \quad j = 2(p+q+1)+2. \end{cases}.$$

In addition to this, it holds that

$$\begin{aligned} &\left|\mathfrak{m}_{\tau,j}(z_t,\theta'',\zeta'',\alpha',\zeta')-\mathfrak{m}_{\tau,j}(z_t,\theta''_0,\zeta_0,\alpha'_0,\zeta_0)\right|^r \\ &\leqslant \tau\left|z_{t,j}(\theta'',\zeta'',\alpha',\zeta')-z_{t,j}(\theta'',\zeta'',\alpha',\zeta')\right|^r \qquad (44)\\ &+\left|z_{t,j}(\theta'',\zeta'',\alpha',\zeta')\mathbb{1}\left\{u_t \leqslant \theta''^{\top}z_t(\alpha',\zeta',\zeta'')\right\}-z_{t,j}(\theta''_0,\zeta_0,\alpha'_0,\zeta_0)\mathbb{1}\left\{u_t\leqslant\theta_0^{\top}z_t(\alpha'_0,\zeta_0,\zeta_0)\right\}\right|^r. \qquad (45)\end{aligned}$$

We start by expanding equation (44) and bounding each term individually:

$$\begin{aligned}\tau\mathbb{E}\left|z_{t,j}(\theta'',\zeta'',\alpha',\zeta')-z_{t,j}(\theta''_0,\zeta_0,\alpha'_0,\zeta_0)\right|^r &\leqslant \tau\mathbb{E}\left|z_{t,j}(\theta'',\zeta'',\alpha',\zeta')-z_{t,j}(\theta''_0,\zeta'',\alpha',\zeta')\right|^r \qquad (46)\\ &+\tau\mathbb{E}\left|z_{t,j}(\theta''_0,\zeta'',\alpha',\zeta')-z_{t,j}(\theta''_0,\zeta'',\alpha'_0,\zeta')\right|^r \qquad (47)\\ &+\tau\mathbb{E}\left|z_{t,j}(\theta''_0,\zeta'',\alpha'_0,\zeta')-z_{t,j}(\theta''_0,\zeta_0,\alpha'_0,\zeta')\right|^r \qquad (48)\\ &+\tau\mathbb{E}\left|z_{t,j}(\theta''_0,\zeta_0,\alpha'_0,\zeta')-z_{t,j}(\theta''_0,\zeta_0,\alpha'_0,\zeta_0)\right|^r. \qquad (49)\end{aligned}$$

Given that $\|\theta''-\theta_0\| < \delta$, a bound for the first term (46) can be obtained by noting that we have finite $(2+\delta)$ moments of u_t (for all t) and finite derivatives of the transition function G by Assumptions 1 and 6 such that $\tau^r\mathbb{E}\left|(\theta''^{\Delta}-\theta_0^{\Delta})^{\top}z_{t,j}(\alpha',\zeta')\frac{\partial G(\zeta'',\eta'')}{\partial\cdot}\right|^r \leqslant K_{1,1,j}\delta^r$. Since $z_{t,j}$ is linear in u_t, the same bound also applies to the expectation of the supremum of the absolute value. For the second term (47) if $\|\alpha'-\alpha_0\| < \delta$, a bound denoted by $K_{1,2,j}\delta^r$ follows immediately from the linearity of $z_{t,j}(\theta'',\zeta'',\alpha',\zeta')$ with respect to α, finite second moments of u_t and Assumption 1. For the remaining terms (48) and (49), with the transition parameters $\|\zeta'' - \zeta_0\| < \delta$ and $\|\zeta' - \zeta_0\| < \delta$, the differentiability of G and the Lipschitz continuity and boundedness of $\frac{\partial G}{\partial \zeta}$ (both stated in Assumption 1), along with previous arguments imply that their respective suprema and the expectations thereof can also be bounded by $K_{1,3,j}\delta^r$ and $K_{1,4,j}\delta^r$, respectively.

Putting all these terms together we get the bound with constant $K_{1,j} = 4 \sup_l K_{1,l,j}$

$$\tau^r \mathbb{E} \sup_{(\theta'', \zeta'', \alpha', \zeta') \in \mathfrak{U}_\delta((\theta''_0, \zeta_0, \alpha'_0, \zeta_0))} \left| z_{t,j}(\theta'', \zeta'' \alpha', \zeta') - z_{t,j}(\theta''_0, \zeta_0, \alpha'_0, \zeta_0) \right|^r \leqslant K_{1,j}\delta^r. \tag{50}$$

Returning to the original inequality (44)-(45), for the second term (45) we note that

$$\begin{aligned}
&\left| z_{t,j}(\theta'', \zeta'', \alpha', \zeta') \mathbb{1}\left\{ u_t \leqslant \theta''^T z_t(\alpha', \zeta', \zeta'') \right\} - z_{t,j}(\theta''_0, \zeta_0, \alpha'_0, \zeta_0) \mathbb{1}\left\{ u_t \leqslant \theta_0^T z_t(\alpha'_0, \zeta_0, \zeta_0) \right\} \right|^r \\
&\leqslant \left| z_{t,j}(\theta'', \zeta'', \alpha', \zeta') \left(\mathbb{1}\left\{ u_t \leqslant \theta''^T z_t(\alpha', \zeta', \zeta'') \right\} - \mathbb{1}\left\{ u_t \leqslant \theta_0^T z_t(\alpha'_0, \zeta_0, \zeta_0) \right\} \right) \right|^r \qquad (51) \\
&+ \left| \left(z_{t,j}(\theta'', \zeta'', \alpha', \zeta') - z_{t,j}(\theta''_0, \zeta_0, \alpha'_0, \zeta_0) \right) \mathbb{1}\left\{ u_t \leqslant \theta_0^T z_t(\alpha'_0, \zeta_0, \zeta_0) \right\} \right|^r. \qquad (52)
\end{aligned}$$

While the bound of the expectation of term (52) follows from (50), we need to take care of equation (51). Let $z'_{t,j} = z_{t,j}(\theta'', \zeta'', \alpha', \zeta')$. Taking expectations in neighbourhoods of the true parameter, it follows

$$\begin{aligned}
&\mathbb{E} \left| z'_{t,j} \left(\mathbb{1}\left\{ u_t \leqslant \theta''^T z_t(\alpha', \zeta', \zeta'') \right\} - \mathbb{1}\left\{ u_t \leqslant \theta_0^T z_t(\alpha'_0, \zeta_0, \zeta_0) \right\} \right) \right|^r \\
&\quad \leqslant \Big\{ \mathbb{E} \left| z'_{t,j} \right|^r \left(\mathbb{1}\left\{ u_t \leqslant \theta''^T z_t(\alpha', \zeta', \zeta'') \right\} - \mathbb{1}\left\{ u_t \leqslant \theta_0^T z_t(\alpha', \zeta', \zeta'') \right\} \right) \\
&\qquad + \mathbb{E} \left| z'_{t,j} \right|^r \left(\mathbb{1}\left\{ u_t \leqslant \theta_0^T z_t(\alpha', \zeta', \zeta'') \right\} - \mathbb{1}\left\{ u_t \leqslant \theta_0^T z_t(\alpha'_0, \zeta', \zeta'') \right\} \right) \\
&\qquad + 2\mathbb{E} \left| z'_{t,j} \right|^r \left(\mathbb{1}\left\{ u_t \leqslant \theta_0^T z_t(\alpha'_0, \zeta', \zeta'') \right\} - \mathbb{1}\left\{ u_t \leqslant \theta_0^T z_t(\alpha'_0, \zeta', \zeta_0) \right\} \right) \\
&\qquad + 2\mathbb{E} \left| z'_{t,j} \right|^r \left(\mathbb{1}\left\{ u_t \leqslant \theta_0^T z_t(\alpha'_0, \zeta', \zeta_0) \right\} - \mathbb{1}\left\{ u_t \leqslant \theta_0^T z_t(\alpha'_0, \zeta_0, \zeta_0) \right\} \right) \Big\} \\
&\quad \leqslant \mathbb{E} \Big\{ \left| z'_{t,j} \right|^r f_{u_t|\mathcal{F}_{t-1}} (\tilde{\phi}_1)\, z_t(\alpha', \zeta', \zeta'')^T (\theta'' - \theta''_0) \qquad (53) \\
&\qquad + \left| z'_{t,j} \right|^r f_{u_t|\mathcal{F}_{t-1}} (\tilde{\phi}_2) \left(G(\zeta'') \sum_{j=1}^{q} \theta^{I}_{p+j+1} L^j z^m_t(\zeta') + (1 - G(\zeta'')) \sum_{j=1}^{q} \theta^{II}_{p+j+1} L^j z^m_t(\zeta') \right)^T (\alpha' - \alpha'_0) \qquad (54) \\
&\qquad + 2 \left| z'_{t,j} \right|^r f_{u_t|\mathcal{F}_{t-1}} (\tilde{\phi}_3) \left(z_t(\alpha'_0, \zeta', \zeta'')^T \theta''^{\Delta}_0 \frac{\partial G(\zeta_0)}{\partial \zeta} \right)^T (\zeta'' - \zeta_0) \qquad (55) \\
&\qquad + 2 \left| z'_{t,j} \right|^r f_{u_t|\mathcal{F}_{t-1}} (\tilde{\phi}_4) \left(G(\zeta'') \sum_{j=1}^{q} \theta^{I}_{p+j+1} L^j z^m_t(\zeta')^T \alpha^{\Delta} \frac{\partial L^j G_t(\zeta_0)}{\partial \zeta} \right. \\
&\qquad\qquad \left. + (1 - G(\zeta'')) \sum_{j=1}^{q} \theta^{II}_{p+j+1} L^j z^m_t(\zeta')^T \alpha^{\Delta} \frac{\partial L^j G_t(\zeta_0)}{\partial \zeta} \right)^T (\zeta' - \zeta_0) \Big\}, \qquad (56)
\end{aligned}$$

where the first inequality follows from the triangle inequality and the fact that $\omega \mapsto \mathbb{1}\{u_t \leqslant \omega\}$ is monotone in ω, which is in turn linear in θ and α. In addition to this, $\mathbb{1} \circ G$ is monotone in the first parameter (location) of ζ, namely ζ, and piece-wise monotone – increasing over half of the domain and decreasing over the other half – in its second parameter η (scale). For the second inequality, we apply the law of iterated expectations and the mean value theorem for which we require the density $f_{u_t|\mathcal{F}_{t-1}}$ to exist and to be bounded (Assumption 4). The variables $\tilde{\phi}_j$ for $j \in \mathcal{I}_{1,4}$ refer to the elements of small neighbourhoods of the respective parameters at which we applied the mean value theorem. While the existence and boundedness of the density and the finiteness of $(2+\delta)$ moments of $z_{t,j}$ (Assumption 5) are sufficient for the terms (53) and (54) not to diverge, the final two terms (55) and (56) additionally require the bound on $\frac{\partial G_t(\zeta)}{\partial \zeta}$ (Assumption 1). Then for any

$(\theta'', \zeta'', \alpha', \zeta)$ in a neighbourhood of their true counterparts $\mathfrak{U}_\delta = \mathfrak{U}_\delta(\theta_0'', \zeta_0, \alpha_0', \zeta_0)$, i.e. where $\|\theta'' - \theta_0\| < \delta$, $\|\zeta'' - \zeta_0\| < \delta$, $\|\alpha' - \alpha_0\| < \delta$ and $\|\zeta' - \zeta_0\| < \delta$, there exists a $K_{2,j} > 0$ such that the second term in equation (45) can be bounded by

$$\mathbb{E} \sup_{(\theta'', \zeta'', \alpha', \zeta') \in \mathfrak{U}_\delta} \left| z_{t,j}(\theta'', \zeta'', \alpha', \zeta') \mathbb{1}\left\{u_t \leqslant \theta''^{\mathsf{T}} z_t(\alpha', \zeta', \zeta'')\right\} - z_{t,j}(\theta_0'', \zeta_0, \alpha_0', \zeta_0) \mathbb{1}\left\{u_t \leqslant \theta_0^{\mathsf{T}} z_t(\alpha_0', \zeta_0, \zeta_0)\right\}\right|^r,$$

which can then be bounded by $K_{2,j}\delta^r$. Subsequently, by Lemma 4.2 in Chen (2008) the following expression holds:

$$\sup_{(\theta,\alpha) \in \mathfrak{U}_\delta(\theta_0(\tau), \alpha_0)} \|g(\theta, \alpha) - g(\theta_0(\tau), \alpha_0) - g_n(\theta, \alpha) + g_n(\theta_0(\tau), \alpha_0)\| = o_p\left(n^{-1/2}\right).$$

Thus, (37) reduces to $\left\|g\left(\widehat{\theta}_n(\tau), \widehat{\alpha}_n\right) - g(\theta_0(\tau), \alpha_0) + g_n\left(\widehat{\theta}_n(\tau), \widehat{\alpha}_n\right) - g_n(\theta_0(\tau), \alpha_0)\right\| = o_p\left(n^{-\frac{1}{2}}\right)$, whereas for (38) we have by definition $\left\|g_n\left(\widehat{\theta}_n(\tau), \widehat{\alpha}_n\right)\right\| = o_p\left(n^{-\frac{1}{2}}\right)$ and we immediately get from equation (35):

$$\lambda_{\min}(\Gamma_{\theta,0}) \left\|\left(\widehat{\theta}_n(\tau) - \theta_0(\tau)\right)\right\| \leqslant \left\|g\left(\widehat{\theta}_n(\tau), \alpha_0\right)\right\| = \mathcal{O}_p\left(n^{-\frac{1}{2}}\right),$$

which completes the proof. □

Corollary 1. *Under Assumptions 1-5, the following linearisation holds as $n \to \infty$:*

$$g_n\left(\widehat{\theta}_n(\tau), \widehat{\alpha}_n\right) - g_n(\theta_0(\tau), \alpha_0) - \Gamma_{\theta,0}(\widehat{\theta}_n(\tau) - \theta_0(\tau)) - \Gamma_{\alpha,0}(\widehat{\alpha}_n - \alpha_0) = o_p\left(n^{-\frac{1}{2}}\right). \tag{57}$$

Proof. By adding and subtracting, equation (57) can be rewritten as:

$$\begin{aligned}
&g_n\left(\widehat{\theta}_n(\tau), \widehat{\alpha}_n\right) - g_n(\theta_0(\tau), \alpha_0) - \Gamma_{\theta,0}(\widehat{\theta}_n(\tau) - \theta_0(\tau)) - \Gamma_{\alpha,0}(\widehat{\alpha}_n - \alpha_0) \\
&\qquad = g_n\left(\widehat{\theta}_n(\tau), \widehat{\alpha}_n\right) - g_n(\theta_0(\tau), \alpha_0) - \Gamma_{\theta,0}(\widehat{\theta}_n(\tau) - \theta_0(\tau)) - \Gamma_{\alpha,0}(\widehat{\alpha}_n - \alpha_0) \\
&\qquad + g\left(\widehat{\theta}_n(\tau), \widehat{\alpha}_n\right) - g\left(\widehat{\theta}_n(\tau), \widehat{\alpha}_n\right) + g(\theta_0(\tau), \alpha_0) - g(\theta_0(\tau), \alpha_0) \\
&\qquad + g\left(\widehat{\theta}_n(\tau), \alpha_0\right) - g\left(\widehat{\theta}_n(\tau), \alpha_0\right) \\
&\qquad + \Gamma_{\alpha,n}\left(\widehat{\theta}_n(\tau), \alpha_0\right)(\widehat{\alpha}_n - \alpha_0) - \Gamma_{\alpha,n}\left(\widehat{\theta}_n(\tau), \alpha_0\right)(\widehat{\alpha}_n - \alpha_0)
\end{aligned}$$

Again taking norms, rearranging the terms on the right hand side, using the triangle inequality, the following bound is obtained as $n \to \infty$:

$$\begin{aligned}
&\left\|g_n\left(\widehat{\theta}_n(\tau), \widehat{\alpha}_n\right) - g_n(\theta_0(\tau), \alpha_0) - \Gamma_{\theta,0}(\widehat{\theta}_n(\tau) - \theta_0(\tau)) - \Gamma_{\alpha,0}(\widehat{\alpha}_n - \alpha_0)\right\| \\
&\leqslant \left\|g_n\left(\widehat{\theta}_n(\tau), \widehat{\alpha}_n\right) - g_n(\theta_0(\tau), \alpha_0) - \left(g\left(\widehat{\theta}_n(\tau), \widehat{\alpha}_n\right) - g(\theta_0(\tau), \alpha_0)\right)\right\| \\
&+ \left\|g\left(\widehat{\theta}_n(\tau), \widehat{\alpha}_n\right) - g\left(\widehat{\theta}_n(\tau), \alpha_0\right)\right) - \Gamma_{\alpha,n}\left(\widehat{\theta}_n(\tau), \alpha_0\right)(\widehat{\alpha}_n - \alpha_0)\right\| \\
&+ \left\|g\left(\widehat{\theta}_n(\tau), \alpha_0\right) - g(\theta_0(\tau), \alpha_0) - \Gamma_{\theta,0}\left(\widehat{\theta}_n(\tau) - \theta_0(\tau)\right)\right\| \\
&+ \left\|\Gamma_{\alpha,n}\left(\widehat{\theta}_n(\tau), \alpha_0\right)(\widehat{\alpha}_n - \alpha_0) - \Gamma_{\alpha,0}(\widehat{\alpha}_n - \alpha_0)\right\| = o_p\left(n^{-\frac{1}{2}}\right),
\end{aligned}$$

where we use stochastic equicontinuity verified in the previous lemma for the first term and reason along the lines of (40), (41), (42) using Lipschitz continuity of $\Gamma_{\alpha,0}$ and $\Gamma_{\theta,0}$ as well as the law of large numbers for $\Gamma_{\alpha,n}$ and $\sqrt{n}$-consistency of $\widehat{\theta}_n(\tau)$ for the remaining terms. □

Proof of Theorem 4: The first order condition $g_n(\theta(\tau), \widehat{\alpha}_n) = 0$ is solved by $\widehat{\theta}_n(\tau)$ so that

$$0 = g_n\left(\widehat{\theta}_n(\tau), \widehat{\alpha}_n\right) = g_n(\theta_0, \boldsymbol{\alpha}_0) + \Gamma_{\theta,0}(\widehat{\theta}_n(\tau) - \theta_0(\tau)) + \Gamma_{\alpha,0}(\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0) + o_p\left(n^{-\frac{1}{2}}\right),$$

using the linearisation from Corollary 1. Since $\Gamma_{\theta,0}$ has full rank by Assumption 9, by pre-multiplying $\sqrt{n}$, an asymptotic representation of the second stage estimator is obtained for $n \to \infty$

$$\sqrt{n}\left(\widehat{\theta}_n(\tau) - \theta_0(\tau)\right) = -\boldsymbol{\Gamma}_{\theta,0}^{-1}\left[\sqrt{n} g_n(\theta_0(\tau), \boldsymbol{\alpha}_0)) + \boldsymbol{\Gamma}_{\alpha,0}\sqrt{n}(\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0)\right] + o_p(1).$$

Finally, we apply the α-mixing central limit theorem (Theorem 18.5.3 in Ibragimov & Linnik, 1971), for which the $(2+\delta)$ moments of the data has to exist and mixing coefficients have to satisfy $\beta_s \to 0$ and $\sum_{s=1}^{\infty} \beta_s^{\delta/(2+\delta)} < +\infty$. These conditions are guaranteed by Assumption 5. After stacking the two summands in the last equation, we can therefore write as $n \to \infty$

$$\begin{aligned}
\sqrt{n}\left(\widehat{\theta}_n(\tau) - \theta_0(\tau)\right) &= \Gamma_{\theta,0}^{-1}\left[\mathbf{I}_{2(p+q+1)}, \Gamma_{\alpha,m,0}\right]\sqrt{n}\begin{bmatrix} g_n(\theta_0(\tau), \boldsymbol{\alpha}_0) \\ \widehat{\boldsymbol{\alpha}_n^{\mathrm{I}}} - \boldsymbol{\alpha}_0^{\mathrm{I}} \\ \widehat{\boldsymbol{\alpha}_n^{\mathrm{II}}} - \boldsymbol{\alpha}_0^{\mathrm{II}} \\ \widehat{\zeta_n} - \zeta_0 \end{bmatrix} + o_P(1) \\
&\approx \frac{1}{\sqrt{n}}\Gamma_{\theta,0}^{-1}\left[\mathbf{I}_{2(p+q+1)}, \frac{\Gamma_{\alpha,m,0}\mathbf{D}_n^{-1}}{\sum_{k=1}^{K} s_k q_k^2}\right]\sum_{t=m+1}^{T}\begin{bmatrix} \begin{bmatrix} \boldsymbol{z}_t(\boldsymbol{\alpha}_0)G_t(\zeta_0) \\ \boldsymbol{z}_t(\boldsymbol{\alpha}_0)(1 - G_t(\zeta_0)) \\ \theta_0^{\Delta T}\boldsymbol{z}_t(\boldsymbol{\alpha}_0)\frac{\partial G_t(\zeta_0)}{\partial \zeta} \end{bmatrix}\left(\mathbb{1}\left\{u_t \leqslant F_{u_t|\mathcal{F}_{t-1}}^{-1}(\tau)\right\} - \tau\right) \\ \begin{bmatrix} G_t(\zeta_0)\,\boldsymbol{z}_t^m \\ (1 - G_t(\zeta_0))\boldsymbol{z}_t^m \\ \boldsymbol{\alpha}_0^{\Delta}\boldsymbol{z}_t^m\frac{\partial G_t(\zeta_0)}{\partial \zeta} \end{bmatrix}\sum_{k=1}^{K} q_k\left(\mathbb{1}\left\{u_t \leqslant F_{u_t|\mathcal{F}_{t-1}}^{-1}(\tau_k)\right\} - \tau_k\right) \end{bmatrix} \\
&\rightsquigarrow \mathcal{N}\left(0, \lim_{n\to\infty}\Gamma_{\theta,0}^{-1}\mathbb{E}\left[\mathbf{M}_t\Xi^{\tau}\mathbf{M}_t^{T}\right]\Gamma_{\theta,0}^{-1}\right),
\end{aligned}$$

where we use independence of the innovations ε_t and $\boldsymbol{z}_t^m$ and $\boldsymbol{z}_t(\boldsymbol{\alpha}_0)$, respectively. The matrices Ξ^{τ}, $\mathbf{M}_t$, and $\Gamma_{\theta,0}$ are defined in Assumption 9, which also postulates the existence of the asymptotic variance matrix. □

A.2 Algorithm for Estimation

Algorithm A.2.1: Two-stage estimation procedure

$(l_{\min}^1, \widehat{\alpha}_n^{\mathrm{I}}, \widehat{\alpha}_n^{\mathrm{II}}, \widehat{\zeta}_n, \widehat{\eta}_n) \leftarrow (\infty, 0, 0, 0, 0)$
for all $\eta \in \{\eta_1, \ldots, \eta_{k_n}\}$ **do**
  Define $G(\xi_t, \zeta, \eta)$, using $\xi_t \leftarrow \xi(\boldsymbol{z}_t)$ for given scale η as a function of ζ
  Estimate $\widehat{\alpha}_{n,k_\eta}^{\mathrm{I}}, \widehat{\alpha}_{n,k_\eta}^{\mathrm{II}}, \widehat{q}_{k_\eta}, \widehat{\zeta}_{k_\eta}$ and obtain loss $l_{k_\eta}^1$ according to (9)
   by (smoothed) composite quantile regression
  if $l_{k_\eta}^1 \leqslant l_{\min}^1$ **then**
   $(l_{\min}^1, \widehat{\alpha}_n^{\mathrm{I}}, \widehat{\alpha}_n^{\mathrm{I}}, \widehat{\zeta}_n, \widehat{\eta}_n) \leftarrow (l_{k_\eta}^1, \widehat{\alpha}_{n,k_\eta}^{\mathrm{I}}, \widehat{\alpha}_{n,k_\eta}^{\mathrm{II}}, \widehat{\zeta}_{k_\eta}, \eta)$
  end if
end for

Calculate $\sigma_t(\widehat{\alpha}_n)$ according to equation (10) using $\widehat{\alpha}_n = (\widehat{\alpha}_n^{IT}, \widehat{\alpha}_n^{IIT}, \widehat{\zeta}_n, \widehat{\eta}_n)^T$
Construct $z_t(\widehat{\alpha}_n) = (\sigma_{t-1}(\widehat{\alpha}_n), \ldots, \sigma_{t-p}(\widehat{\alpha}_n), |u_{t-1}|, \ldots, |u_{t-q}|)^T$
$(l^2_{\min}, \widehat{\theta}^I_n, \widehat{\theta}^{II}_n, \widehat{\zeta}_n, \widehat{\eta}_n) \leftarrow (\infty, 0, 0, 0, 0)$
for all $(\zeta, \eta) \in \{\zeta_1, \ldots \zeta_{k_\zeta}\} \times \{\eta_1, \ldots, \eta_{k_\eta}\} \cap 3$ **do**
Calculate $G(\xi_t, \zeta, \eta)$ using $\xi_t \leftarrow \xi(z_t)$ for given location ζ and scale η
Estimate $\widehat{\theta}^I_{n,k_{\zeta,\eta}}, \widehat{\theta}^{II}_{n,k_{\zeta,\eta}}$ and obtain loss $l^2_{k_{\zeta,\eta}}$ according to (11)
by linear (inequality constrained) quantile regression
if $l^2_{k_{\zeta,\eta}} \leqslant l^2_{\min}$ **then**
$(l^2_{\min}, \widehat{\theta}^I_n, \widehat{\theta}^{II}_n, \widehat{\zeta}_n, \widehat{\eta}_n) \leftarrow (l^2_{k_{\zeta,\eta}}, \widehat{\theta}^I_{n,k_{\zeta,\eta}}, \widehat{\theta}^{II}_{n,k_{\zeta,\eta}}, \zeta, \eta)$
end if
end for
Set the final estimate to $\widehat{\theta}_n = (\widehat{\theta}_n^{IT}, \widehat{\theta}_n^{IIT}, \widehat{\zeta}_n, \widehat{\eta}_n)^T$.

A.3 Sufficient conditions for absolute regularity

Let us recall existing results regarding the (nonlinear) GARCH processes. Meitz & Saikkonen (2008) derive Theorems 1 and 2 stating sufficient conditions for the geometric ergodicity of Markov processes following various nonlinear GARCH models and implying their stationarity and absolute regularity (see Meitz & Saikkonen, 2008, Section 2.2 and Theorem 3). We limit ourselves for simplicity to the models of order 1 here with the transition variable being the lagged dependent variable.

First, various regularity assumptions (Meitz & Saikkonen, 2008, Assumption 2) have to hold so that the volatility process defined by (3) or a GARCH model have some basic properties such as irreducibility and aperiodicity. These regularity assumptions are however satisfied in our and GARCH models due to the imposed assumptions (i.e., Assumption 3) and the model definitions, implying the boundedness of volatility function on compact subsets of the support, its positivity on the compact subsets of the support, and its monotonicity and differentiability in the transition variable (Assumption 1). More importantly, Theorems 1–3 of Meitz & Saikkonen (2008) require that the volatility is bounded by a linear function of the past volatility and its slope has a finite expectation smaller than 1. Denoting the volatility in (1) by σ_t for the sake of conciseness and recalling that all parameter values are assumed to be non-negative, the standard GARCH(1,1) volatility process in model (1) is for example bounded by

$$\sigma_t^2 = \beta_0 + \beta_1 \sigma_{t-1}^2 + \gamma_1 u_{t-1}^2 \leqslant \beta_0 + [\beta_1 + \gamma_1 \varepsilon_{t-1}^2]\sigma_{t-1}^2.$$

Assuming that $E(\beta_1 + \gamma_1 \varepsilon_{t-1}^2) < 1$ holds, that is, $\beta_1 + \gamma_1 < 1$ if the variance of ε_{t-1} is normalised to 1, Meitz & Saikkonen (2008, Theorems 1–3) with $V(\sigma) = 1 + \sigma$ imply that the GARCH(1,1) volatility process is V-geometrically ergodic and the GARCH(1,1) process itself is geometrically ergodic and β-mixing.

Under the above mentioned regularity assumptions, Theorems 1–3 of Meitz & Saikkonen (2008) allow us to directly obtain the same results also for the single-regime model (1)–(2), where for the first-order model

$$\sigma_t = \beta_0 + \beta_1 \sigma_{t-1} + \gamma_1 |u_{t-1}| \leqslant \beta_0 + [\beta_1 + \gamma_1 |\varepsilon_{t-1}|]\sigma_{t-1}$$

and it has to hold that $E(\beta_1 + \gamma_1 |\varepsilon_{t-1}|) < 1$ and $\beta_1 + \gamma_1 < 1$ if $E|\varepsilon_{t-1}|$ is normalised to 1. Finally, a similar condition can be found for the general model (8) (again of the first order), where

$$\begin{aligned}
\sigma_t &= G_t(\zeta)[\beta_0^I + \beta_1^I \sigma_{t-1} + \gamma_1^I |u_{t-1}|] + [1 - G_t(\zeta)][\beta_0^{II} + \beta_1^{II} \sigma_{t-1} + \gamma_1^{II} |u_{t-1}|] \\
&\leqslant \max\{\beta_0^I, \beta_0^{II}\} + [\max\{\beta_1^I, \beta_1^{II}\} + \max\{\gamma_1^I, \gamma_2^{II}\} |\varepsilon_{t-1}|] \sigma_{t-1}
\end{aligned}$$

since $G_t(\zeta) \in [0,1]$. To apply Meitz & Saikkonen (2008, Theorems 1–3), it has to hold that $E([\max\{\beta_1^I, \beta_1^{II}\} + \max\{\gamma_1^I, \gamma_2^{II}\}|\varepsilon_{t-1}|]) < 1$, that is, $\max\{\beta_1^I, \beta_1^{II}\} + \max\{\gamma_1^I, \gamma_2^{II}\} < 1$ if $E|\varepsilon_{t-1}| = 1$.

The theorems above can be also extended to model (3), but we have to formulate the volatility as a multivariate process as the volatility σ_t depends on volatilities σ_t^I and σ_t^{II} of each regime:

$$\begin{aligned}
\sigma_t^I &= \beta_0^I + \beta_1^I \sigma_{t-1}^I + \gamma_1^I |u_{t-1}| \\
\sigma_t^{II} &= \beta_0^{II} + \beta_1^{II} \sigma_{t-1}^{II} + \gamma_1^{II} |u_{t-1}| \\
\sigma_t &= G_t(\zeta)[\beta_0^I + \beta_1^I \sigma_{t-1}^I + \gamma_1^I |u_{t-1}|] + [1 - G_t(\zeta)][\beta_0^{II} + \beta_1^{II} \sigma_{t-1}^{II} + \gamma_1^{II} |u_{t-1}|].
\end{aligned}$$

Defining $V(\sigma^I, \sigma^{II}, \sigma) = 1 + \sigma$, the proof of Meitz & Saikkonen (2008, Theorem 1) applies to $(\sigma_t^I, \sigma_t^{II}, \sigma_t)$ using $V(\sigma^I, \sigma^{II}, \sigma)$ since by definition (3),

$$\begin{aligned}
\sigma_t &= G_t(\zeta)[\beta_0^I + \beta_1^I \sigma_{t-1}^I + \gamma_1^I |u_{t-1}|] + [1 - G_t(\zeta)][\beta_0^{II} + \beta_1^{II} \sigma_{t-1}^{II} + \gamma_1^{II} |u_{t-1}|] \\
&= G_t(\zeta)[\beta_0^I + \beta_1^I \sigma_{t-1}^I + \gamma_1^I |\varepsilon_{t-1}| \sigma_{t-1}] + [1 - G_t(\zeta)][\beta_0^{II} + \beta_1^{II} \sigma_{t-1}^{II} + \gamma_1^{II} |\varepsilon_{t-1}| \sigma_{t-1}] \\
&\leqslant \max\{\beta_0^I, \beta_0^{II}\} + [\max\{\beta_1^I, \beta_1^{II}\} + \max\{\gamma_1^I, \gamma_2^{II}\} |\varepsilon_{t-1}|] \sigma_{t-1}.
\end{aligned}$$

Hence, if $E[\max\{\beta_1^I, \beta_1^{II}\} + \max\{\gamma_1^I, \gamma_2^{II}\}|\varepsilon_{t-1}|] < 1$, that is, $\max\{\beta_1^I, \beta_1^{II}\} + \max\{\gamma_1^I, \gamma_2^{II}\} < 1$ after normalisation of $E|\varepsilon_{t-1}|$ to 1, $(\sigma_t^I, \sigma_t^{II}, \sigma_t)$ is V-geometrically ergodic, and following Meitz & Saikkonen (2008, Theorems 2–3), we can obtain the geometric ergodicity and β-mixing properties for the ANST-GARCH process.

A.4 Simulation Results

By default, the estimation is performed for time series of length $n = 1000$, the number of simulations per experiment is $s = 100$, the composite quantile regression employs by default $k = 9$ quantiles for $\tau \in [0.05, 0.25] \cup [0.75, 0.95]$, the truncation parameter for the ARCH approximation is set to $m = \lceil \frac{5}{2} n^{\frac{1}{4}} \rceil$ and the grid size is $(k_\zeta, k_\eta) = (30, 30)$. The true global parameter vector for both processes is chosen to be $\theta_0 = (\beta_0^I, \beta_1^I, \gamma_1^I, \beta_0^{II}, \beta_1^{II}, \gamma_1^{II})_0^T = (0.50, 0.15, 0.60, 0.25, 0.30, 0.15)^T$ and the location-scale parameter pair equals $\zeta_0 = (\zeta, \eta)_0^T = (0.00, 0.2)^T$. While β_0^I and β_0^{II} are only determining the unconditional variances of the respective regimes, we chose γ_1^I and γ_1^{II} in a way that is consistent with findings in the two regime conditional heteroscedasticity literature (Gonzales-Rivera, 1998; Lubrano, 2001; Wago, 2004; Khemiri, 2011). Unfortunately, the findings on regime-specific parameter values for β_1^I and β_1^{II} are rather limited and there is also no clear link to their single regime counterparts. Thus coming up with a sensible prior is somewhat *ad hoc*. We approached this by choosing their values in a way that generates both a higher and a lower persistence regime. Unreported simulations show that different DGPs work similarly well, although, perhaps unsurprisingly, numerical stability deteriorates as one of the regimes' processes becomes close to being integrated.

If not stated otherwise, we will assume the innovations to be standard normally distributed: $\varepsilon_t \sim N(0,1)$. When running simulations using different innovation distributions, in order to ensure comparability, their variances will always be normalised to one. This implies that there is one high and one low variance regime with unconditional variances, defined by $\beta_0^r/(1-\beta_1^r-\gamma_1^r)$ for $r \in \{I, II\}$, of 2 and 0.45, respectively. All of the presented results use a specification with the logistic function G_{logistic}. However, unreported simulations confirmed that the GACQ estimation is insensitive to the misspecification of the transition function (e.g., if the logistic transition function is used while the true underlying model follows the linear or threshold function). Finally, note that we have to restrict the grid for both location and scale. We introduce the data-driven criterion ensuring that location satisfies $\zeta \in [\underline{\zeta}, \bar{\zeta}]$ with unconditional sample quantiles $\underline{\zeta} = \widehat{F}_{u_t}^{-1}(0.1)$ and $\bar{\zeta} = \widehat{F}_{u_t}^{-1}(0.9)$. Similarly, the scale is restricted to $\eta \in [\underline{\eta}, \bar{\eta}(\underline{\zeta}, \bar{\zeta})]$ with fixed $\underline{\eta} = 0.1$ and $\bar{\eta}(\underline{\zeta}, \bar{\zeta}) = \left[\log(0.1^{-1} - 1)(0.5\bar{\zeta} - 0.5\underline{\zeta})\right]^{-1}$. The latter bound represents the inverse of the logistic function with respect to the scale evaluated at 0.1 and the location at the centre of the considered location grid.

To evaluate the procedures, we report the biases and root mean squared errors (RMSE) of all estimates. As the focus of the quantile regression modelling is on the estimation of quantiles such as Value at Risk rather than parameters, the performance is measured by the mean (absolute) prediction error averaged over the sample, denoted as M(A)PE, absolute one-period-ahead out-of-sample forecast errors (MAFE) and by the coverage ratio, each of them referring to the estimated 5% Value at Risk. Note that the coverage (ratio) is defined as the proportion of observations falling below the estimated Value at Risk and should thus be close to $\tau = 0.05$ for the 5% Value at Risk. It should be mentioned that while coverage, MPE, MAPE and MAFE are reported in the Bias column for the purpose of a tidy exposition, their values represent the mean deviations from the value 0.0, which corresponds to the perfect fit of the model: for example, coverage value 0 represents the exactly correct coverage level 0.05 and MAPE value 0 would represent the exact fit. The RMSE of these quantities additionally depict their corresponding Monte Carlo standard deviations. We will use these metrics to compare different estimators with each other as well as the impact of different features of the data generating process on prediction and forecasting.

Table S.5: The bias and RMSE of GACQ for different sample sizes n.

	$n = 1000$		$n = 2000$		$n = 4000$	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_0^{I}	0.1797	0.4120	0.1223	0.2489	0.0749	0.2123
β_1^{I}	-0.0129	0.3559	0.0065	0.2779	-0.0068	0.1179
γ_1^{I}	-0.1196	0.3270	-0.0653	0.2276	-0.0228	0.1404
β_0^{II}	0.0319	0.1650	0.0337	0.1231	0.0236	0.1006
β_1^{II}	0.0263	0.2274	0.0272	0.1701	0.0187	0.1021
γ_1^{II}	0.0050	0.1310	0.0011	0.0984	0.0150	0.0891
ζ	0.3798	0.4923	0.3252	0.4703	0.4595	0.5632
$\zeta(\tau)$	0.1063	0.3730	0.1149	0.3138	0.0541	0.2547
η	0.0870	0.1341	0.0627	0.1170	0.0743	0.1229
$\eta(\tau)$	-0.0622	0.0867	-0.0327	0.0807	-0.0353	0.0826
MPE	0.0082	0.0482	0.0027	0.0447	0.0032	0.0203
MAPE	0.1259	0.1310	0.0974	0.1004	0.0682	0.0700
MAFE	0.1129	0.1420	0.0848	0.1182	0.0530	0.0718
coverage	0.0008	0.0013	0.0004	0.0007	0.0001	0.0003

Table S.6: The bias and RMSE of GACQ as a function of the order of the first-stage ARCH(m) approximation with $m = \lceil cn^{1/4} \rceil$.

	$c = 2.0$		$c = 2.5$		$c = 3.0$		$c = 3.5$	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_0^{I}	0.2070	0.4318	0.1727	0.3901	0.2041	0.4427	0.1837	0.4149
β_1^{I}	-0.0031	0.3767	0.0173	0.3590	0.0037	0.3643	-0.0286	0.3606
γ_1^{I}	-0.1409	0.3316	-0.1264	0.3310	-0.1441	0.3568	-0.1154	0.3110
β_0^{II}	0.0377	0.1666	0.0295	0.1714	0.0317	0.1714	0.0205	0.1777
β_1^{II}	0.0091	0.2168	0.0444	0.3932	0.0075	0.2022	0.0181	0.1984
γ_1^{II}	0.0047	0.1319	0.0085	0.1354	0.0070	0.1265	0.0123	0.1359
ζ	0.3779	0.4898	0.3939	0.4912	0.4618	0.5484	0.4818	0.5685
$\zeta(\tau)$	0.1380	0.3901	0.0955	0.3617	0.1198	0.3849	0.0855	0.3644
η	0.0573	0.1175	0.0412	0.0970	0.0740	0.1319	0.0635	0.1197
$\eta(\tau)$	-0.0531	0.0864	-0.0608	0.0882	-0.0503	0.0860	-0.0605	0.0888
MPE	0.0104	0.0504	0.0117	0.0507	0.0128	0.0516	0.0104	0.0516
MAPE	0.1273	0.1321	0.1248	0.1297	0.1270	0.1324	0.1270	0.1316
MAFE	0.1009	0.1307	0.1074	0.1409	0.1070	0.1502	0.1091	0.1473
coverage	0.0008	0.0013	0.0004	0.0013	0.0008	0.0014	0.0007	0.0014

Table S.7: The bias and RMSE of GACQ if quantiles $\tau \in (0.5 - \delta/2, 0.5 + \delta/2)$ are not used in estimation.

	$\delta = 0.15$		$\delta = 0.20$		$\delta = 0.30$		$\delta = 0.50$	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_0^{I}	0.1748	0.4086	0.1897	0.4217	0.1816	0.4322	0.2021	0.4270
β_1^{I}	0.0106	0.3652	0.0133	0.3825	-0.0035	0.3550	-0.0078	0.3897
γ_1^{I}	-0.1275	0.3335	-0.1374	0.3503	-0.1258	0.3465	-0.1385	0.3509
β_0^{II}	0.0281	0.1610	0.0415	0.1655	0.0260	0.1611	0.0391	0.1589
β_1^{II}	0.0281	0.2159	0.0283	0.2271	0.0474	0.2331	0.0191	0.2095
γ_1^{II}	0.0040	0.1213	-0.0024	0.1125	-0.0030	0.1222	0.0004	0.1206
ζ	0.3641	0.4925	0.3341	0.4811	0.3311	0.4334	0.3259	0.4628
$\zeta(\tau)$	0.1073	0.3647	0.1211	0.3811	0.1017	0.3608	0.1348	0.3727
η	0.0639	0.1226	0.0717	0.1210	0.0748	0.1223	0.0606	0.1148
$\eta(\tau)$	-0.0532	0.0863	-0.0581	0.0849	-0.0571	0.0853	-0.0536	0.0868
MPE	0.0116	0.0496	0.0105	0.0518	0.0092	0.0484	0.0111	0.0498
MAPE	0.1228	0.1275	0.1262	0.1310	0.1260	0.1308	0.1263	0.1318
MAFE	0.1042	0.1392	0.1123	0.1408	0.1134	0.1426	0.0955	0.1208
coverage	0.0008	0.0014	0.0008	0.0014	0.0009	0.0014	0.0006	0.0012

Table S.8: The bias and RMSE of the GACQ and GARCH estimators in the case of normally distributed errors.

	GACQ		GARCH-N		GARCH-t	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_0^{I}	0.1797	0.4120	0.0343	0.0986	-0.1489	0.2789
β_1^{I}	-0.0129	0.3559	-0.1351	0.1410	-0.1020	0.1405
γ_1^{I}	-0.1196	0.3270	-0.1507	0.1752	-0.0145	0.2500
β_0^{II}	0.0319	0.1650	-0.0820	0.1007	0.0180	0.2037
β_1^{II}	0.0263	0.2274	-0.0190	0.0991	-0.0938	0.1928
γ_1^{II}	0.0050	0.1310	-0.0571	0.1094	0.1146	0.2908
ζ	0.3798	0.4923	0.1248	0.3144	0.4136	0.6883
$\zeta(\tau)$	0.1063	0.3730	0.1248	0.3144	0.4136	0.6883
η	0.0870	0.1342	-0.1478	0.1593	73247	248720
$\eta(\tau)$	-0.0622	0.0867	-0.1478	0.1593	73247	248720
MPE	0.0082	0.0482	0.0771	0.0861	-0.2142	0.3402
MAPE	0.1259	0.1310	0.1461	0.1568	0.3320	0.3877
MAFE	0.1129	0.1420	0.2939	0.4278	0.4325	0.5591
coverage	0.0008	0.0013	0.0071	0.0094	-0.0150	0.0219

Table S.9: The bias and RMSE of the GACQ and GARCH estimators in the case of Student's t_4 distributed errors.

	GACQ		GARCH-N		GARCH-t	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_0^I	0.2086	0.5754	0.0209	0.1717	0.0298	0.1314
β_1^I	0.0547	0.6786	-0.0723	0.1494	-0.1037	0.1293
γ_1^I	-0.1722	0.4184	-0.2043	0.2504	-0.1763	0.2143
β_0^{II}	0.0455	0.2513	-0.0824	0.1638	-0.1066	0.1271
β_1^{II}	0.0429	0.3827	-0.0279	0.1677	-0.0270	0.1181
γ_1^{II}	0.0024	0.1858	-0.0749	0.1085	-0.0753	0.0939
ζ	0.3625	0.4689	0.1240	0.3008	0.1386	0.2587
$\zeta(\tau)$	0.1940	0.4330	0.1240	0.3008	0.1386	0.2587
η	0.0304	0.0781	-0.1453	0.1657	-0.1267	0.1560
$\eta(\tau)$	-0.0845	0.0927	-0.1453	0.1657	-0.1267	0.1560
MPE	0.0046	0.0596	-0.0325	0.0676	-0.3239	0.3368
MAPE	0.1536	0.1585	0.1409	0.1517	0.3271	0.3397
MAFE	0.1286	0.1798	0.3164	0.5886	0.4561	0.6462
coverage	0.0008	0.0013	-0.0044	0.0081	-0.0254	0.0259

Table S.10: The bias and RMSE of the GACQ and GARCH estimators in the case of errors following the re-centered Type 1 Gumbel distribution.

	GACQ		GARCH-N		GARCH-t	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_0^I	0.1368	0.2631	-0.0183	0.0286	-0.0113	0.0694
β_1^I	-0.0295	0.2081	-0.0857	0.1405	-0.1261	0.2264
γ_1^I	-0.0838	0.1921	-0.0076	0.0573	0.0045	0.0550
β_0^{II}	0.0615	0.1230	0.1304	0.1903	0.1162	0.1783
β_1^{II}	-0.0365	0.1255	-0.1858	0.2224	-0.1296	0.1966
γ_1^{II}	-0.0025	0.1153	0.0949	0.1826	0.0571	0.1329
ζ	0.2723	0.4382	-0.1837	0.2432	0.4859	2.4011
$\zeta(\tau)$	0.1346	0.2908	-0.1837	0.2432	0.4859	2.4011
η	0.0575	0.0967	-0.0476	0.1115	-0.0615	0.0898
$\eta(\tau)$	-0.0418	0.0751	-0.0476	0.1115	-0.0615	0.0898
MPE	0.0108	0.0279	0.0351	0.0414	-0.2860	0.2961
MAPE	0.0727	0.0750	0.1045	0.1161	0.2905	0.2999
MAFE	0.0611	0.0836	0.1261	0.1815	0.5775	0.6107
coverage	0.0006	0.0014	0.0017	0.0064	-0.0263	0.0269

Table S.11: The bias and RMSE of the GACQ and GARCH estimators in the case of normally and Student distributed errors with 2.5% outliers.

	GACQ: $\varepsilon \sim \mathcal{N}$		GARCH-N: $\varepsilon \sim \mathcal{N}$		GACQ: $\varepsilon \sim t(4)$		GARCH-t: $\varepsilon \sim t(4)$	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_0^{I}	0.2656	0.6630	-0.3962	0.4538	0.2758	0.7464	-0.0975	0.2817
β_1^{I}	0.0358	0.4642	0.0647	0.3166	0.1187	0.6819	-0.0614	0.1429
γ_1^{I}	-0.0749	0.2867	0.0804	0.3435	-0.0615	0.3810	-0.0911	0.2344
β_0^{II}	0.0668	0.2337	0.4326	0.4719	0.1801	0.5841	0.1139	0.2651
β_1^{II}	0.0315	0.2780	-0.0313	0.2028	0.0659	0.5414	-0.0089	0.1635
γ_1^{II}	-0.0010	0.1379	-0.1096	0.1308	-0.0290	0.1701	-0.0689	0.0972
ζ	0.1545	0.5263	1.0059	1.0553	0.1686	0.5461	0.3365	0.5907
$\zeta(\tau)$	0.1832	0.5171	1.0059	1.0553	0.1599	0.5316	0.3365	0.5907
η	0.0678	0.1588	-0.1480	0.3122	0.0262	0.1183	-0.0419	0.5081
$\eta(\tau)$	-0.0445	0.0929	-0.1480	0.3122	-0.0805	0.0939	-0.0419	0.5081
MPE	-0.1023	0.1291	-175.09	1008.6	-0.1528	0.1940	-0.9166	0.9684
MAPE	0.2180	0.2311	175.22	1008.6	0.2749	0.2933	0.9182	0.9696
MAFE	0.2175	0.4203	0.5545	0.6154	0.2791	0.5217	0.9664	1.1980
coverage	0.0007	0.0013	-0.0144	0.0164	0.0009	0.0014	-0.0269	0.0273